Affine and Quadratic Models for Volatility and Interest Rates Markets

Dissertation
for the Faculty of Economics, Business Administration and Information Technology of the University of Zurich

to achieve the title of
Doctor of Philosophy
in Banking & Finance

presented by

Elise Gourier

from Pontailler-sur-Saône, France

approved in July 2013 at the request of

Prof. Dr. Markus Leippold
Prof. Dr. Josef Teichmann

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich hereby authorizes the printing of this Doctoral Thesis, without thereby giving any opinion on the views contained therein.

Zurich, July 2013.

The Chairman of the Doctoral Committee: Prof. Dr. Dieter Pfaff.

Acknowledgements

I want to express my gratitude to Prof. Dr. Markus Leippold, my thesis supervisor, for his support and advice, and for the great work atmosphere I could benefit from. Furthermore he gave me the chance to prepare and teach the Financial Engineering lecture together with him and Chris Bardgett, which was a very enriching experience. I would like to further acknowledge the support of my co-supervisor Prof. Dr. Josef Teichmann, whose valuable inputs on my conference and seminar presentations have improved my work.

I am also very grateful to Prof. Dr. Loriano Mancini and Prof. Dr. Damir Filipović who gave me the great opportunity to work with them on a very interesting topic. I learnt a lot from them both on a technical and on a more human level, and admire their rigor and continued enthusiasm for research. I would like to thank them for the time they dedicated to me, the ideas they shared and their patience when difficulties arose.

I also thank Chris Bardgett who is my coauthor and with whom I worked long hours on the paper, sometimes late at night when we were under time pressure. He was always available, even during his stay in London, and working together was both fruitful and pleasant.

I feel privileged to have worked on publications together with Markus, Loriano, Damir and Chris, and hope that we can pursue our collaboration in the future.

I would also like to thank Prof. Dr. Walter Farkas and Prof. Dr. Marc Paolella for always being encouraging and for the career advice they gave me.

Furthermore, I had the chance to be in a wonderful chair and enjoyed very much my time as a PhD student. Special thanks goes to Felix Matthys for his patience and to my colleagues and friends Lujing Su, Caroline Oehri, Nikola Vasiljevic and Meriton Ibraimi, for always being ready to help and for all the good times we had together.

Finally I wish to express deep gratitude to my family and friends who has always supported and encouraged me throughout the ups and downs.

Zürich, June 2013 Elise Gourier

Contents

Part I

Introduction

Introduction and Summary of Research Results

Volatility and interest rates share many features. In the model of Black and Scholes (1973) for stock price returns, they are both considered constant. However, modelling them as realistically as possible and investigating their features have become one of the main goals of mathematical finance, producing a sophisticated growing body of literature. Three approaches are predominant in interest rates and volatility modelling. The first one focuses on the instantaneous processes. Both short rate and volatility are commonly represented by a square root process, which ensures their positivity and reflects their stationarity. Because this approach has difficulties to accurately represent the term structure of interest rates and volatility, an alternative is to model the corresponding forward processes. In interest rates modelling this idea was introduced by Heath, Jarrow, and Morton (1992), and more recently Buehler (2006) developed a class of similar models for the forward variance. Unfortunately, the instantaneous interest rate and forward rates (respectively instantaneous/forward variances) are not directly observable quantities in financial markets, making the choice of a particular model and statistical inference complicated. The third class of models, market models, were developed to overcome this issue by directly modelling observable quantities, such as Libor rates. These models are very popular among practitioners working with interest rates derivatives; indeed, in their simplest form where the forward Libor (alternatively swap rate) is modelled by a Geometric Brownian Motion, closed-form expressions are available for options, caps and floors (alternatively swaptions).

This doctoral thesis entitled "Affine and Quadratic Models for Volatility and Interest Rates Markets" comprises three papers which investigate the ability of affine models to represent different features of volatility and interest rates, and price derivatives on these underlyings. Interest rates derivatives are by far the largest derivatives market in the world. With a notional amount outstanding of USD 418 billion in 2008,[1] they represent more than 70% of the total amount outstanding in the global OTC derivatives market. On the other side, derivatives on volatility have attracted growing attention in the last decade. According to financial press (e.g., Gangahar (2006)), variance swaps have become the preferred tool used by market participants to bet on and/or hedge against volatility movements. Furthermore, since their introduction in 2006, options on the volatility index VIX have gained increasing popularity, and are with options on the S&P 500 among the most liquid worldwide with a daily average volume of 391,992 traded contracts (783,768 on the S&P 500) in 2011.

The recent financial crisis has had a tremendous impact on both interest rates and volatility markets. It uncovered the impact of counterparty and liquidity risks on some interest rates spreads and gave

[1] According to the report from the Bank of International Settlements published in May 2009.

rise to phenomena that had never been observed in the past. Because rates with different tenors were affected by different levels of risk, some essential classical arbitrage relationships were violated and new modelling methodologies were needed. On the other side, because of the leverage effect, volatility reacted very strongly to the crisis as well. The VIX index increased from about 20% up to about 80% in less than a month end of 2008, variance swap rates exhibited a similar peak and smiles of volatility shifted upwards. These sudden movements led to a number of analyses that pointed to the presence of jumps in the interest rates and volatility processes. Under the historical measure, these jumps would be justified by the large upward movements in the trajectories of the processes, while under the risk-neutral measure they could explain the steep smiles of volatility which are typically observed for short-maturity options.

This thesis mainly answers three questions. First, what information can we infer on volatility from S&P 500 and VIX underlying levels and option prices using affine processes? Second, do quadratic models improve on affine models in representing the variance term structure and how can this be exploited in a trading strategy? Third, how can affine models be used to build a Libor Market Model which is consistent with the stylized facts of interest rates and reflects the spreads that appeared during the crisis? Each paper is presented in a separate chapter, organized as follows.

In the first Chapter *Inferring volatility dynamics and risk-premia from the S&P 500 and VIX markets*, we use a large dataset of S&P 500 and VIX index and option prices with wide ranges of maturities and moneynesses, and analyze the empirical performance of affine jump-diffusion models for S&P 500 returns to jointly represent underlyings' and derivatives' prices. Based on the affine relationship of the VIX squared with respect to the latent factors, we extend the Fourier Cosine Expansion to efficiently price VIX derivatives. We build an Auxiliary Particle Filter which sources the information contained in both indices and derivatives' prices over time and investigate the behavior of the filtered latent processes. We analyze the out-of-sample performance of sub-models depending on which products and markets are considered in the in-sample estimation procedure. We find that a stochastic central tendency is needed to better represent the tails of the returns' distribution and the term structure of the smiles of volatility on both S&P 500 and VIX markets. Furthermore, jumps in returns and volatility help reproduce the tail of the variance distribution. Finally, we investigate and compare the information that the underlying levels and options contain on latent factors and risk premia.

In the second Chapter *Quadratic Variance Swap Models*, we introduce a novel class of term structure models for variance swaps. The multivariate state variable follows a diffusion process characterized by a quadratic diffusion function. The variance swap curve is quadratic in the state variable, and available in closed form in terms of a linear ordinary differential equation, greatly facilitating empirical analysis. Various goodness-of- fit tests show that quadratic models fit variance swaps on the S&P 500 remarkably well and outperform nested specifications, including popular affine models. An empirical study of a dynamic optimal portfolio in variance swaps and the S&P 500 reveals the versatility of quadratic models, and the economic value of variance swaps.

In the third Chapter *Libor Market Model: How to account for the Crisis?*, we build a new model

for Libor rates, which accounts for the stylized effects that appeared during the financial crisis in the dynamics of rates with different tenor structures. Since liquidity and counterparty risks associated to Libor rates depend on the length of the borrowing/lending period, the model is based on a multi-curve approach and reflects the discrepancies which have appeared between rates that used to chase one and another. We define specific dynamics for every Libor rate, depending on its tenor structure, and use Lévy processes as drivers to accommodate for jumps. We provide closed-form expressions in the general setup for the prices of basic interest-rate derivatives. Moreover, we investigate for special cases of Lévy processes, the role of the different parameters on the spread between forward and FRA rates.

REFERENCES

Black, F., and M. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.

Buehler, H., 2006, "Consistent Variance Curve Models," *Finance and Stochastics*, 10, 178–203.

Gangahar, A., 2006, "Volatility Becomes an Asset Class," *Financial Times*, May 23.

Heath, D., R. Jarrow, and A. Morton, 1992, "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Evaluation," *Econometrica*, 60, 77–105.

Part II

Research Papers

Inferring volatility dynamics and risk-premia from the S&P 500 and VIX markets

Chris Bardgett, Elise Gourier and Markus Leippold

I have presented this paper at:

- ETH-UZH Finance and Mathematics Doctoral Seminar, November 2012, Zürich, Switzerland.
- 5th International Conference of the ERCIM WG on Computing & Statistics, December 2012, Oviedo, Spain.

Abstract

We use a large dataset of S&P 500 and VIX index and option prices with wide ranges of maturities and moneynesses, and analyze the empirical performance of affine jump-diffusion models for S&P 500 returns to jointly represent underlyings' and derivatives' prices. Based on the affine relationship of the VIX squared with respect to the latent factors, we extend the Fourier Cosine Expansion to efficiently price VIX derivatives. We build an Auxiliary Particle Filter which sources the information contained in both indices and derivatives' prices over time and investigate the behavior of the filtered latent processes. We analyze the out-of-sample performance of sub-models depending on which products and markets are considered in the in-sample estimation procedure. We find that a stochastic central tendency is needed to better represent the tails of the returns' distribution and the term structure of the smiles of volatility on both S&P 500 and VIX markets. Furthermore, jumps in returns and volatility help reproduce the tail of the variance distribution. Finally, we investigate and compare the information that the underlying levels and options contain on latent factors and risk premia.

1 Introduction

One of the central questions addressed by research in empirical option pricing is the determination of asset returns dynamics. Ideally, in addition to reproducing asset prices and prices of derivatives traded on a given day on the market, they should also recreate the joint evolution of these prices over time. Under the historical measure $\mathbb{P}$ the time series of asset returns provide valuable information on the main characteristics of returns dynamics (assuming stationarity and ergodicity). On the other hand option prices on this asset help to specify its dynamics under the risk neutral measure $\mathbb{Q}$. Indeed, the result of Breeden and Litzenberger (1978) states that the observation of vanilla option prices with maturity T for a continuum of strikes entirely determines the $\mathbb{Q}$ distribution of this asset at the future time T. Even though we do not observe prices for arbitrary strikes in practice, the S&P 500 index has many strikes traded liquidly. Whereas a large part of the literature on asset pricing either focuses on the time-series properties of returns under the historical measure or proposes models to accurately capture the stylized facts of option prices, the study of the link between both measures has recently captured more attention. The change of measure from $\mathbb{P}$ to $\mathbb{Q}$ is achieved through an appropriate specification of risk premia, which can be interpreted as compensations for the risks that investors take when buying an asset. While there is a large amount of research articles and surveys on the equity risk premium, the study of the variance dynamics and variance risk premium is more recent. Prominent examples include Bates (1996, 2000, 2003), Chernov and Ghysels (2000), Jackwerth (2000), Pan (2002), Jones (2003), Eraker (2004), Aït-Sahalia and Kimmel (2007), Broadie, Chernov, and Johannes (2007), Carr and Wu (2009), Todorov (2010) and Wu (2011). However, the components of risk premia, in particular when jumps are involved, are usually found hard to estimate and statistically insignificant. One reason for this is that the estimation of risk premia requires a large amount of returns and options data and therefore powerful computational tools to extract the relevant information. In fact, because of this computational burden, most research does not make use of the whole cross section of options and considerably reduces the amount of information. As available sources of information have grown tremendously since the introduction of the volatility index VIX as well as VIX derivatives, the need for efficient computational algorithms arises.

The VIX index has been constructed to approximate non-parametrically the expected future realized volatility of the S&P 500 returns over the next 30 days. The VIX is not directly tradeable but it is possible to trade VIX futures and options. The options started trading in 2006 and have been a growing business at a dramatic speed ever since. They now represent a much larger market than VIX futures. By definition, the VIX index is linked to the dynamics of the S&P 500 index returns and this makes VIX and S&P 500 options both ideal to infer these dynamics. Fortunately, the S&P 500 and VIX options markets are among the most liquid worldwide with a daily average volume of 783,768 and 391,992 contracts traded per day in 2011, and therefore represent a trustworthy source of information. Including more information on volatility and its evolution over time is essential to better specify and understand the dynamics of volatility. Subsequently, we will use interchangeably S&P 500 and SPX, which is its ticker symbol.

Our contribution is the following: We develop an algorithm that uses time series of returns and

derivatives on both the SPX and VIX markets to investigate the historical and risk-neutral dynamics of the SPX returns. First of all, we base our analysis on a time series of cross sectional data of options with a wide range of moneynesses and maturities. We emphasize that we have kept liquid deep out-of-the money options in our dataset in contrast to most of the literature, so as to keep valuable information about the tails of S&P 500 returns. These options are considered to be essential in the estimation of the jump structure of their underlying under the risk-neutral measure, i.e., of the jumps of S&P 500 and VIX indices. Indeed, the steepness of the S&P 500 smile and the high volatilities for short maturities puts is considered to be a strong indication of jumps in the returns. Similarly, the positive skewness of VIX implied volatilities and high volatilities of deep out-of-the-money calls can be explained by the presence of positive jumps in the VIX. Therefore, the cross section of options is required to infer possible jumps under the risk-neutral measure and justifies why we have decided to use the whole cross section of S&P 500 and VIX options. Second, including derivatives on both the S&P 500 and the VIX indices allows us to make better inference on model parameters and risk premia dynamics, which are consistent with both markets. Up to now and to our knowledge, extracting information from both SPX and VIX derivatives markets has not been done and therefore provides new valuable insight into the dynamics of asset returns and volatility.

We model the S&P 500 returns using the affine framework of Duffie, Pan, and Singleton (2000). This structure allows us to price S&P 500 and VIX derivatives in semi-closed form and is essential to carry out the analysis of returns and volatility dynamics using such a large dataset of options. However, we point out and reduce the limitations of one-factor affine models by advocating a stochastic level of reversion in the volatility dynamics, which is a key ingredient in order to consistently accommodate for the time series of both markets. The flexibility of this model helps to investigate how many factors are needed to reproduce the times series features of the data, and whether jumps should be incorporated.

Estimating the dynamics using such an extremely large dataset of options on the two markets and for a long time series requires computationally efficient techniques that can easily deal with the complicated features of the model, in particular state-dependent jumps. To achieve this goal, we extend the Cosine method introduced by Fang and Oosterlee (2008) for S&P 500 options to price VIX options and adapt the Auxiliary Particle Filter of Pitt and Shephard (1999) to filter out unobservable processes over time and their jumps. Sequential Monte-Carlo techniques have been recently used for limited datasets, but most papers restrict their dataset of options to near at-the-money options. Furthermore, we want to stress that our estimation methodology consists in a single step using the times series of indices and options together. In particular, we do not estimate the model under the historical measure and then fix parameters to estimate the risk-neutral dynamics. Our approach increases the computational complexity but ensures a consistent estimation of the historical and pricing measures, which is essential to estimate reliably risk premia.

Among the outputs of the Auxiliary Particle Filter are the filtered trajectories of the latent processes. We investigate their behavior, compare them for some sub-models and using different estimation datasets. This analysis allows us to draw conclusions regarding the usefulness of jumps and the number of factors needed to represent volatility. Furthermore, we investigate the information content

of the underlying levels and of the options on each market, compare them and provide a discussion on risk premia.

The paper is organized as follows. In Section 2, we explain how our paper fits into the existing literature. We then conduct a preliminary data analysis in Section 3, and highlight some differences between the S&P 500 and the VIX option markets. In Section 4 we present the affine two-factor model with jumps that we use in the estimation. We describe the risk premium specification and derive the pricing formula for the VIX squared as well as VIX and S&P 500 options. In Section 5 we discuss the joint estimation to one single day of data as well as the Auxiliary Particle Filter that we use to calibrate the model to a time-series of cross-sectional data. Finally, in Sections 6 and 7 we summarize the results of the daily and time-series estimations and present our findings. Section 8 concludes.

2 Related literature

Our work builds on an extensive body of research that analyzes which features are needed for a model to provide a realistic representation of equity underlying and derivatives prices. While the end of the twentieth century has been characterized by a fast growing literature on equity option pricing, the financial crisis has recently drawn more attention to the need to better understand and model equity volatility. Since the introduction of the volatility index VIX in 1993 and its derivatives (from 2004 onwards), the direct modeling of volatility and the pricing of its derivatives has been the focus of numerous papers. We refer among others to Whaley (1993), Grünbichler and Longstaff (1996), Detemple and Osakwe (2000), Bergomi (2004, 2005, 2008), Sepp (2008a,b), Bergomi (2009), Lian and Zhu (2011), Drimus and Farkas (2013) and Mencía and Sentana (2013). An important conclusion of this literature is that sharp increases in the variance dynamics are necessary to reproduce the positive skewness of VIX options' implied volatilities. In particular, many articles point out that this could be achieved by having positive jumps in the variance. In particular, Christoffersen, Jacobs, and Mimouni (2010) demonstrate via Q-Q plots that using a square-root model without jumps for the variance is not in line with empirical properties of the data. Using realized variance taken at high-frequency as a proxy for the integrated variance, they show that the empirical realized volatility is not Gaussian as the continuous square-root model posits. Jumps allow the distribution of the integrated volatility to be fatter-tailed and therefore represent better the data. Todorov (2010) tests for jumps in the VIX index and finds strong evidence supporting this assumption. He also tests for co-jumps in S&P 500 returns and in the VIX and finds striking evidence for them. He finally finds that 63% of the co-jump variation in the sample studied is due to the combination of negative jumps in the returns and positive jumps in the volatility. Eraker, Johannes and Polson (2000) show that using jumps in the volatility process significantly improves the fit of returns. Finally, as mentioned in Eraker (2004), continuous volatility or variance processes are not able to explain the unusually large volatility before and after the crash of 1987. The specification of jumps is furthermore of importance. Bates (1996), Pan (2002) and Eraker (2004) argue in favor of using state-dependent jumps in returns, which is

intuitively appealing as jumps tend to occur more frequently when volatility increases. Using variance swaps, Aït-Sahalia, Karaman, and Mancini (2012) found that the state dependent intensity of jumps was a desirable model feature. However, evidence supporting this choice is mixed. Indeed, Bates (2000) finds that state dependent intensities lead to strong misspecification and Eraker (2004) finds that it does not significantly improve the option prices fit. Broadie, Chernov, and Johannes (2007) and Johannes, Polson, and Stroud (2009) use a constant intensity of jumps.

Another concern of volatility modeling relates to the number of factors that should be used. While adding an additional factor to the Heston model increases the complexity, it has indeed been shown that two factors are needed to provide an accurate description of the volatility dynamics (see, e.g., Andersen, Benzoni, and Lund (2002), Alizadeh, Brandt, and Diebold (2002), Chernov, Gallant, Ghysels, and Tauchen (2003), Todorov (2010), Kaeck and Alexander (2012), Bates (2012) and Mencía and Sentana (2013)).

Several papers have been published in the last years aiming to reconcile the cross-sectional information of the S&P 500 and the VIX derivatives markets by modeling them jointly. Gatheral (2008) pointed out first that even though the Heston model performs fairly well to price S&P 500 options, it totally fails to price VIX options. Figure 4 shows that modeling the instantaneous volatility as a square root process leads to a VIX smile decreasing with moneyness, which is the opposite of what is observed in practice. Therefore the volatility density implied by VIX options has more mass at high volatility and less mass at lower volatility levels than the Chi-Square density of the Heston model. Some studies are going in the direction of non-affine models (e.g., Jones (2003), Aït-Sahalia and Kimmel (2007), Christoffersen, Jacobs, and Mimouni (2010), Ferriani and Pastorello (2012), Durham (2012), Kaeck and Alexander (2012)). However tractability remains an issue that is of crucial importance when it comes to calibrating a model to a long time series containing hundreds of options each day.

Among the recent papers that attempted to reproduce simultaneously the smiles of volatility of S&P 500 and VIX options are Chung, Tsai, Wang, and Wenig (2011), Cont and Kokholm (2011), Song and Xiu (2012), Papanicolaou and Sircar (2012) and Bayer, Gatheral, and Karlsmark (2013). We build on this literature by considering extensions of the Heston model that remain in the affine framework, but add more flexibility to the specifications used in the above mentioned papers. Our model is a special case of the general affine framework developed by Duffie, Pan, and Singleton (2000) but includes as sub-cases the usual extensions of the Heston model encountered in the literature, for example Bates (2000), Eraker (2004) and Sepp (2008a).

However most if not all of the papers that consider S&P 500 and VIX options in their calibration exercise have restricted their analysis to a static one-day estimation. Therefore the estimated parameters might exhibit large variations when calibrating the model to different dates. Lindström, Ströjby, Brodén, Wiktorsson, and Holst (2008) show that the estimated parameters are not stable over time and therefore cannot be used to infer time series properties of returns and risk premia. In the last decade, powerful algorithms have been developed to estimate non-linear models with non-Gaussian innovations in a time-consistent manner.

Time-consistent estimation methods have been used so far to calibrate models to index returns and

options. For example, Pan (2002) uses a tailored version of the Generalized Methods of Moments to estimate the Bates model using a time series of S&P 500 and options (two per day). Eraker (2004) relies on Markov Chain Monte Carlo methods to estimate risk premia for jumps in returns and volatility also using returns and options (around three per day). Broadie, Chernov, and Johannes (2007) were the first to consider the whole cross section of option prices on the S&P 500. To reduce the computational burden, they fix some of the parameters by taking values from previous estimations of the time series of returns and minimize a least square type distance between market and model option implied volatilities. They find that the time series provided evidence that volatility jumps, which coincides with the literature that appeared later on VIX option pricing. With a particle filter, Johannes, Polson, and Stroud (2009) investigate whether the time-series of returns of the S&P 500 are consistent with information embedded in option prices. Their options sample is limited to one option per day. They find some inconsistencies that they attribute to either a wrong specification of risk premia or a lack of flexibility of the model. They conclude that their results might be explained by the introduction of a time varying level of reversion for the volatility. Christoffersen, Jacobs, and Mimouni (2010) apply a Maximum Likelihood Importance Sampling technique on returns and a separate Non-linear Least-Squares Important Sampling estimation to option prices to compare the accuracy of models in reproducing returns and option prices. However, as underlined in Ferriani and Pastorello (2012), most papers filtering information from option prices rely on one option per day or a very limited set of options. This is computationally less intensive but ignores a large part of the information present on the market. Ferriani and Pastorello (2012) have used part of the cross section of options and the time series of log-returns in the filtering problem. They do not consider jumps in the volatility but study different non-affine models. They conclude that significant improvement could be brought into these models by incorporating jumps or regime switching in the volatility dynamics. Finally, in a working paper Duan and Yeh (2011) use a filter on the S&P 500 returns together with the VIX index to infer the dynamics of returns and volatility. However, they do not use options data making it impossible to estimate risk premia.

3 Preliminary data analysis

In Figure 1, we plot the joint evolution of the S&P 500 and the VIX index. Their movements are highly negatively correlated, which explains the use of instruments on the VIX to hedge part of the equity risk of a portfolio. Table 1 displays the first four moments of the S&P 500 returns and VIX index levels, over two periods of time. The first period covers from March 2006 until November 2008, i.e., it spans the pre-crisis period as well as the beginning of the crisis. The second period starts in December 2008 and lasts until October 2010. Log-returns on the S&P 500 exhibit negative skewness during the second period considered, and a high kurtosis over both periods, suggesting the presence of rare and large movements. The VIX index exhibits a large skewness and kurtosis in the first period, but in the second period the statistics suggest that the movements are more symmetric, centered about a higher value (29% instead of 20% in the first period).

[Insert Figure 1 here]

[Insert Table 1 here]

We consider closing prices of European options on the S&P 500 from March 1, 2006 to October 29, 2010. The data was obtained from OptionMetrics. The time period of our dataset is restricted by the fact that options on the VIX were introduced in 2006. We also use a dataset of VIX options closing prices on the same time period coming from the data provider DeltaNeutral. This time series includes periods of calm and periods of crisis with extreme events, especially relevant to estimate the presence and magnitude of jumps. In particular, during the financial crisis that started at the beginning of 2007, the VIX index was at its highest peak since its launch.

Both the S&P 500 and VIX options dataset are treated following usual procedures (see Aït-Sahalia and Lo (1998)). In particular, we only consider options with maturity between one week and one year and delete options quotes that where not traded on a given date. We follow two main steps. First, we delete all in-the-money (ITM) options since they are illiquid compared to out-of-the-money (OTM) options. Second, we infer from highly liquid options the Futures price using the at-the-money (ATM) put-call parity. This avoids two issues: Making predictions on future dividends, and using Futures closing prices which are not synchronized with the option closing prices. Hence, we consider that the underlying of the options is the index Futures and not the index itself. At the end, we only work with liquid OTM options for the S&P 500 market and only with liquid call options for the VIX market. Indeed, in the case where the VIX ITM call is not liquid, we use the put-call parity to infer from a more liquid VIX OTM put a liquid VIX ITM call.

These adjustments leave a total of 383,286 OTM S&P 500 options and a total of 43,775 call options on the VIX. This implies a daily average of 327 S&P 500 options and 37 VIX options. The number of S&P 500 options in our dataset on a given date is increasing with time with around 170 options at the beginning of the dataset and around 450 options at the end. For VIX options, the number is increasing substantially, with around 5 options per day at the beginning and around 70 options per day at the end. At the beginning of the sample, there are one or two short maturities (below 6 months) available for VIX options and around 6 maturities for S&P 500 options with approximately 40 options per maturity slice. At the end of the sample, VIX options have around 5 short maturities (less than 6 months) with a bit more than 10 options trading per maturity. For S&P 500 options, around ten maturities are available per day with around 60 options for one-month maturities and 40 options for the one-year slice. The low number of VIX options compared to the number of S&P 500 options is first coming from the fact that the VIX options market started in 2006 and therefore that the overall volume traded is lower but also from the fact that less maturities and less strikes are traded. At the end of our sample, the total VIX options volume per day is about half the total volume of S&P 500 options traded but much fewer strikes are traded for VIX options.

It is important to understand that calculating implied volatilities of VIX options using as underlying the VIX index is incorrect. Indeed, the true underlying of VIX options is the VIX Futures value. This

can intuitively be explained by the fact that a call option at time t with maturity T is an option on volatility on the time interval $[T, T + 30d]$, where $30d$ stands for 30 days. The value VIX_t at time t is related to volatility on the time interval $[t, t + 30d]$ which might not overlap at all with $[T, T + 30d]$. On the contrary, a Futures on the VIX with maturity T is based on the volatility on the time interval $[T, T + 30d]$. This remark is important because traded VIX option prices do not satisfy no-arbitrage relations with respect to VIX index, but rather with respect to the VIX Futures value. In particular, calculating implied volatilities assuming that the underlying is the VIX might lead to volatilities equal to zero, or which simply do not exist. For this reason all implied volatilities are calculated with respect to the Futures price of the VIX. The same is done for S&P 500 options as it eliminates the need to make predictions on futures dividends.

Even though the S&P 500 and VIX markets are related, we want to emphasize that VIX options behave in a completely different way than S&P 500 options. First, S&P 500 and VIX derivatives with the same maturity contain different information. On the one hand, an S&P 500 option with maturity T contains information about the future S&P 500 index level at time T and therefore about the S&P 500 volatility up to T. On the other hand, a VIX option with maturity T embeds information about the VIX at time T and therefore about the S&P 500 volatility between T and $T + 30$ days. Second, the implied volatility smiles backed out from S&P 500 and VIX option prices have very different shapes. Figure 2 displays the S&P 500 and VIX smiles depending on different states of the economy. These implied volatilities (IVs) are computed using the Black-Scholes formula, i.e., backing out the standard deviation of a log-normal distribution for the S&P 500 index (respectively for the VIX index) that are implied by their respective option prices. The VIX IVs are in general substantially higher - ranging from 40% to 200%, with an average IV of around 75% (see Table 2) - than S&P 500 IVs (average IV of around 23%). The implied volatilities are negatively skewed for S&P 500 options, generally decreasing with moneyness as risk-averse investors require a premium for negative states of the economy. In contrast, VIX implied volatilities are positively skewed and increase with moneyness, which can intuitively be explained by the fact that negative returns are often observed together with a rise of volatility (the so-called leverage effect) also corresponding to turbulent states of the economy.

[Insert Figure 2 here]

[Insert Table 2 here]

The difference between these markets is also reflected by other indicators such as the put-call trading ratio: Almost twice as many puts as calls are traded daily in the S&P 500 options market but the situation is reversed in the VIX market where the amount of calls traded daily is almost the double of that of the puts. In fact, one can additionally see in Figure 2 that the log-moneynesses traded for S&P 500 options are mostly negative (which corresponds to out-of-the-money put options) and often positive for VIX options (out-of-the-money calls).

Figure 3 represents the expected forward returns of the underlying S&P 500 index returns from March 1st, 2006 to October 29th, 2010 as implied by prices of S&P 500 options with maturity 1 month.

We use the method described in Bakshi, Kapadia, and Madan (2003) to calculate the moments implied by option prices. The expected forward returns illustrates the variety of market situations that our time series includes. They were almost constant until the end of 2007, equal to a positive value and thus indicating that market participants were expecting a stable income from investing in the index. But from the end of 2007 they exhibit more variation and seem to mean-revert around a negative trend. Suddenly, following the bail out of Lehman Brothers in September 2008, expected forward returns drop and reach -2% beginning of October 2008. Then they gradually come back and stabilize in mid-2009 around a slightly negative level close to -0.2%. In 2010, the sudden increase in the VIX index coincides with a peak of the expected forward returns reaching about -0.8%. It is interesting to compare the VIX index and expected forward returns as implied by S&P 500 options. Indeed, both indicate market expectations over the next month as reflected in index option prices. However volatility provides information on returns through the leverage effect, while the implied expected forward returns are a direct measure of how investors expect returns to behave. They are much more stable in quiet periods and better reflect the different market situations that compose our time-series and that we aim to reproduce with a model.

[Insert Figure 3 here]

Panel B of Figure 3 displays the expected forward returns on the VIX as implied by VIX options. They remain negative throughout the time series, with large peaks that occur simultaneously with the two peaks of the VIX, reaching between -20% and -25%.

4 Model and option pricing

In this section we present the three-factor affine model that we use. This class of models is known to yield semi-closed form expressions for the price of European options on the S&P 500 index. We show that an additional advantage of this model is that the VIX squared can be expressed as an affine function of the variance and of its level of mean reversion. This allows us to use Fourier analysis and derive semi closed-form expressions for the prices of European claims on the VIX as well.

4.1 Model specification

We consider a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ satisfying the usual assumptions, where $\mathbb{P}$ denotes the historical measure. We fix a risk-neutral measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ and denote by $(F_t)_{t\geq 0}$ the forward price[2] of the S&P 500 index and by $Y = (Y_t)_{t\geq 0} = (\log(F_t))_{t\geq 0}$ the returns. The dynamics of Y under $\mathbb{Q}$ are specified as follows:

[2]Assuming that the interest rate r and dividend yield are constant, it does not matter which maturity of the forward we consider because the cash-and-carry relationship between the forward and the spot index ensures that all forwards have the same dynamics (but different initial conditions).

$$dY_t = [-\lambda^Y(v_{t^-}, m_{t^-})(\theta_Z^{(\mathbb{Q})}(1,0,0) - 1) - \frac{1}{2}v_{t^-}]dt + \sqrt{v_{t^-}}dW_t^{Y(\mathbb{Q})} + dJ_t^{Y(\mathbb{Q})} \quad (1)$$

$$dv_t = \kappa_v^{(\mathbb{Q})}(m_{t^-} - v_{t^-})dt + \sigma_v\sqrt{v_{t^-}}dW_t^{v(\mathbb{Q})} + dJ_t^{v(\mathbb{Q})} \quad (2)$$

$$dm_t = \kappa_m^{(\mathbb{Q})}(\theta_m^{(\mathbb{Q})} - m_{t^-})dt + \sigma_m\sqrt{m_{t^-}}dW_t^{m(\mathbb{Q})} + dJ_t^{m(\mathbb{Q})} \quad (3)$$

where W^Y, W^v, W^m are three $\mathbb{Q}$ Brownian motions and

$$d\langle W^Y, W^v\rangle_t = \rho_{Y,v}dt \ ; \ d\langle W^m, W^Y\rangle_t = 0 \ ; \ d\langle W^m, W^v\rangle_t = 0. \quad (4)$$

Our model is a two-factor stochastic volatility model with jumps, which allows the variance process $(v_t)_{t\geq 0}$ of the forward returns to revert towards a stochastic central tendency $(m_t)_{t\geq 0}$. Egloff, Leippold, and Wu (2010) show that this model provides an improvement over the one-factor model in pricing variance swaps. The processes J^Y, J^v, J^m are finite activity jump processes defined by:

$$dJ_t^Y = Z_t^{Y(\mathbb{Q})}dN_t^{Yv} \ ; \ dJ_t^v = Z_t^{v(\mathbb{Q})}dN_t^{Yv} \ ; \ dJ_t^m = Z_t^{m(\mathbb{Q})}dN_t^m. \quad (5)$$

As suggested by the simultaneous peaks in the S&P 500 and VIX index, and in the expected forward returns on both indices, large movements in the equity returns and in the variance are likely to occur at the same time. Therefore we choose, in line with the literature (see, e.g., Cont and Kokholm (2011)) to use the same Poisson process to generate jumps in the asset returns and in the variance process. We also choose the intensity of jumps to be dependent on the level of the factors. Formally, N_t^m and N_t^{Yv} are Poisson processes with respective intensities:

$$\lambda^m(m_{t^-}) = \lambda_0^m + \lambda_1^m m_{t^-} \quad (6)$$

$$\lambda^{Yv}(v_{t^-}, m_{t^-}) = \lambda_0^{Yv} + \lambda_1^{Yv}v_{t^-} + \lambda_2^{Yv}m_{t^-} \quad (7)$$

Moreover, the 3-dimensional process $Z^{(\mathbb{Q})} = (Z^{Y(\mathbb{Q})}, Z^{v(\mathbb{Q})}, Z^{m(\mathbb{Q})})^\top$ (where $.^\top$ denotes the transpose operator) corresponds to the random jump sizes under $\mathbb{Q}$ and we assume that their values taken at two times t and s are independent and identically distributed (i.i.d.) for all $t \neq s$. We assume that jump sizes in the forward returns are normally distributed $\mathcal{N}(\mu_Y^{(\mathbb{Q})}, \sigma_Y^{(\mathbb{Q})})$ and that the jump sizes in the two volatility factors are exponentially distributed with respective means $\nu_v^{(\mathbb{Q})}$ and $\nu_m^{(\mathbb{Q})}$. These jumps sizes are characterized by their joint Laplace transform:

$$\theta_Z^{(\mathbb{Q})}(\phi) = \theta_Z^{(\mathbb{Q})}(\phi_Y, \phi_v, \phi_m) = \mathbb{E}^{\mathbb{Q}}[\exp(\phi^\top Z^{(\mathbb{Q})})], \quad (8)$$

where $\phi \in \mathbb{C}^3$.

4.2 Risk premium specification

We specify the change of measure from the pricing to the historical measure so that the model dynamics keep the same structure under $\mathbb{P}$. The parameters under $\mathbb{P}$ will simply have a superscript referring to the historical measure. Similarly to Broadie, Chernov, and Johannes (2007), we separate the total equity risk premium γ_t into a Brownian contribution which is proportional to the variance level and represents the compensation for the diffusive price risk, and a jump contribution which reflects the compensation for jump risk:

$$\gamma_t = \eta_Y v_{t^-} + \lambda^Y(v_{t^-}, m_{t^-})(\theta_Z^{(\mathbb{P})}(1,0,0) - \theta_Z^{(\mathbb{Q})}(1,0,0)). \tag{9}$$

where $\theta_Z^{(\mathbb{P})}$ denotes the joint Laplace transform of jump sizes under the historical measure $\mathbb{P}$.

As in Pan (2002) and Eraker (2004) we impose the intensity of jumps to be the same under $\mathbb{Q}$ and $\mathbb{P}$.[3]

We define the mean price jump risk premium as the difference between the mean of the jump sizes in returns under $\mathbb{Q}$ and $\mathbb{P}$. Analogously, the volatility of price jump risk premium refers to the difference between the volatility of the jump sizes in returns under $\mathbb{Q}$ and $\mathbb{P}$.[4]

We proceed similarly with the volatility risk premium and decompose it into a diffusive component and a jump component. The diffusive variance risk premium in v is proportional to the current level of variance, with coefficient of proportionality given by:

$$\eta_v = \kappa_v^{(\mathbb{Q})} - \kappa_v^{(\mathbb{P})}. \tag{10}$$

This risk premium should primarily be identified by the term structure of SPX implied volatilites as well as the cross section of VIX implied volatilities. The jump part of the volatility risk premium refers to the difference between the mean of the jump sizes in the variance under $\mathbb{Q}$ and $\mathbb{P}$.

Finally, we introduce a risk premium in the stochastic central tendency, which consists of a diffusive part proportional to the variance of m with coefficient of proportionality given by:

$$\eta_m = \kappa_m^{(\mathbb{Q})} - \kappa_m^{(\mathbb{P})}. \tag{11}$$

The corresponding jump risk premium in m is the difference between the mean of the jump sizes in m under $\mathbb{Q}$ and $\mathbb{P}$. The two latter risk premia should be identified by the cross section and term structure

[3]Pan (2002) argues that introducing different intensities of jumps under the historical and pricing measure introduces a jump-timing risk premium that is very difficult to disentangle from the mean jump risk premium. The consequence of this assumption is that the jump-timing risk premium is artificially incorporated into the mean jump size risk premium.
[4]In the literature σ_Y has sometimes been constrained to be the same under $\mathbb{P}$ and $\mathbb{Q}$ (Bates (1988), Naik and Lee (1990)), but this is not required by absence of arbitrage and we follow Broadie, Chernov, and Johannes (2007) by allowing them to be different. Indeed, they find strong evidence for them to be different and report that this has strong implications for the magnitude of the premium attached to the mean price jump size.

of the VIX implied volatilities as well as the long-term SPX implied volatilities. Therefore, introducing options in our dataset with various moneynesses and maturities is crucial to have meaningful values for these premia.

Finally, no-arbitrage considerations force the volatility of volatilities (σ_v and σ_m) and the correlation between the returns and volatility $\rho_{Y,v}$ to be equal under $\mathbb{P}$ and $\mathbb{Q}$.

4.3 Derivatives pricing

The joint model presented for the S&P 500 and its stochastic volatility implicitly defines a model for the VIX. The VIX index is formally defined in the white paper of the CBOE (2009) and is calculated in practice using a combination of S&P 500 options with maturities adjacent to 30 days. Intuitively, the VIX squared is close to the 30-day expected future realized variance and therefore the value of the VIX index should be close to the 30 day-variance swap on the S&P 500 returns. Demeterfi, Derman, Kamal, and Zou (1999) showed that variance swaps can be partially hedged (and therefore priced) using a combination of vanilla options and this is where the formal definition of the VIX is coming from.

In the following, we do not make the assumption that the VIX is approximately the 30-day realized volatility. Instead, we use its definition as a finite sum of call and put prices that converges (under the assumption that there exists call and put options for all strikes in $\mathbb{R}_+$) to the integral

$$\begin{aligned}\mathrm{VIX}_t^2 &= \frac{2}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\tau}\frac{dF_u}{F_{u^-}} - d(\ln F_u)\right] \\ &= \frac{1}{\tau}\mathbb{E}_t^{\mathbb{Q}}\left[\int_t^{t+\tau} v_u dt + 2\left(e^{Z_u^{Y(\mathbb{Q})}} - 1 - Z_u^{Y(\mathbb{Q})}\right) dN_u^Y\right].\end{aligned}$$

In the affine model we use, the expression of the VIX can be derived and is given by:

Proposition 4.1. *The VIX squared at time t can be written as an affine function of v_t and m_t:*

$$VIX_t^2 = \alpha_{VIX^2} v_t + \beta_{VIX^2} m_t + \gamma_{VIX^2} \tag{12}$$

where the coefficients α_{VIX^2}, β_{VIX^2} and γ_{VIX^2} are known in closed-form.

The coefficients α_{VIX^2}, β_{VIX^2} and γ_{VIX^2} are provided in the Appendix 1.

Due to the affine property of the VIX^2, we have the following result:

Proposition 4.2. *The Laplace transforms of the returns and VIX^2 defined by the model* (1) - (3) *are*

given by

$$\Psi_{VIX_T^2}(t, v_t, m_t; \omega) := \mathbb{E}_t^{\mathbb{Q}}\left[e^{\omega VIX_T^2}\right],$$
$$\Psi_{Y_T}(t, y_t, v_t, m_t; \omega) := \mathbb{E}_t^{\mathbb{Q}}\left[e^{\omega Y_T}\right],$$

are exponential affine in the factor processes:

$$\Psi_{VIX_T^2}(t, v, m; \omega) = e^{\alpha(T-t)+\beta(T-t)\cdot v+\gamma(T-t)\cdot m},$$
$$\Psi_{Y_T}(t, y, v, m; \omega) = e^{\alpha_Y(T-t)+\beta_Y(T-t)\cdot y+\gamma_Y(T-t)\cdot v+\delta_Y(T-t)\cdot m},$$

where α, β, γ, α_Y, β_Y, γ_Y and δ_Y are functions defined on $[0, T]$ by ODEs presented in the Appendix 2. $\omega \in \mathbb{C}$ is chosen so that the expectations above are well defined.

In affine models, option pricing is most efficiently performed using Fourier inversion techniques since we know the Fourier transform of the stochastic processes of interest. To price options on the S&P 500, Fang and Oosterlee (2008) report that the Fourier Cosine Expansion is very efficient and fast compared to other Fourier inversion technique. We use their technique to price S&P 500 options and extend it to incorporate also the pricing of VIX options. This technique is comparable to the inversion performed by Sepp (2008a) but more parsimonious in the number of computational parameters.

Pricing options on the VIX poses technical difficulties that are not encountered when pricing equity options. To understand why it is different, let us write the price of a call option with strike K and maturity T on the VIX at time $t = 0$

$$C(\text{VIX}_0, K, T) = e^{-rT}\int_0^\infty (\sqrt{v} - K)^+ f_{\text{VIX}_T^2}(v)dv, \tag{13}$$

where $f_{\text{VIX}_T^2}$ is the density of the VIX^2 at time $t = T$. We introduce the density of VIX^2 because this is the variable which is affine in our framework (as opposed to working with the VIX).

The square root appearing in the integral as part of the payoff prevents us from using the Fast Fourier Transform of Carr and Madan (1999). For S&P 500 call options the payoff can be written as $(e^y - K)^+$ where y will be the log of the stock price. The fact that we have the exponential e^y allows to interpret this integral as a Fourier transform. To apply the same methodology in the case of VIX derivatives, we would need the log of the VIX to be affine which is incompatible with affine models. This justifies our choice to depart from the standard Fourier pricing techniques.

The basic idea of the method developed by Fang and Oosterlee (2008) is to write the density of the S&P 500 log-returns as a Fourier cosine expansion on a well chosen truncated interval $[a, b]$. This allows them to derive the price of S&P 500 options, we use the same methodology to calculate the price of VIX options.

Theorem 4.1. *Let us consider a European style contingent claim on the VIX index with maturity*

T and payoff $u_{VIX}(VIX^2) = (\sqrt{VIX^2} - K)^+$ (respectively S&P 500 index and payoff $u_{SPX}(e^y)$). Assuming that we are at time $t = t_0 \geq 0$, the price $P_{VIX}(t_0, VIX_0; T)$ (respectively $P_{SPX}(t_0, Y_0; T)$) of the contingent claim is

$$P_{VIX}(t_0, VIX_0; T) = \sum_{n=0}^{N-1} A_n^{VIX^2} U_n^{VIX^2}, \tag{14}$$

respectively

$$P_{SPX}(t_0, Y_0; T) = \sum_{n=0}^{N-1} A_n^{Y} U_n^{Y}. \tag{15}$$

The terms in the sum are defined by:

$$A_n^{VIX^2} = \frac{2}{b_{VIX} - a_{VIX}} Re\left\{ \Psi_{VIX_T^2}\left(VIX_0, \frac{n\pi}{b_{VIX} - a_{VIX}} \right) \exp\left(-ia \frac{n\pi}{b_{VIX} - a_{VIX}} \right) \right\}, \tag{16}$$

$$U_n^{VIX^2} = \int_{a_{VIX}}^{b_{VIX}} \left(\sqrt{v} - K\right)^+ \cos\left(n\pi \frac{v - a_{VIX}}{b_{VIX} - a_{VIX}} \right) dv \tag{17}$$

and similarly

$$A_n^{Y} = \frac{2}{b_Y - a_Y} Re\left\{ \Psi_{Y_T}\left(Y_0, \frac{n\pi}{b_Y - a_Y} \right) \exp\left(-ia \frac{n\pi}{b_Y - a_Y} \right) \right\}, \tag{18}$$

$$U_n^{Y} = \int_{a_Y}^{b_Y} u_{SPX}(e^y) \cos\left(n\pi \frac{y - a_Y}{b_Y - a_Y} \right) dy. \tag{19}$$

We note that the coefficient A_n is computed using Proposition 4.2 and the coefficient U_n is known in closed form and given in Appendix 3.

5 Joint estimation and particle filter

The goal of this section is twofold. First, we explain how we calibrate the nested models (1) - (3) to S&P 500 and VIX options, i.e., estimate the model under the pricing measure using one day of VIX and S&P 500 options. This exercise allows us to show that the $\mathbb{Q}$ dynamics of the model is sufficiently rich to accurately price both S&P 500 and VIX derivatives together, i.e., at any date t we can find a fixed set of parameters which allows the model to price both the VIX options and the S&P 500 options accurately. Second and most importantly, we detail how we have built a time consistent estimation of the models using a time series of S&P 500 and VIX indices together with a time series of S&P 500 and VIX option prices. This means that we estimate both $\mathbb{P}$ and $\mathbb{Q}$ dynamics of the model using the time series of indices and options (i.e., we find one vector of parameters for the whole time series of SPX/VIX spots and SPX/VIX options). The algorithm we use is the Auxiliary Particle Filter, introduced by Pitt and Shephard (1999). It allows to filter out unobserved latent variables, such as

the volatility process or jumps.

From Section 4.1, we recall the $\mathbb{P}$- and $\mathbb{Q}$- parameter vectors:

$$\Theta^{\mathbb{P}} = \{\kappa_v^{(\mathbb{P})}, \kappa_m^{(\mathbb{P})}, \theta_m^{(\mathbb{P})}, \nu_m^{(\mathbb{P})}, \nu_v^{(\mathbb{P})}, \mu_Y^{(\mathbb{P})}, \sigma_Y^{(\mathbb{P})}, \eta_Y\} \tag{20}$$

$$\Theta^{\mathbb{Q}} = \{\kappa_v^{(\mathbb{Q})}, \kappa_m^{(\mathbb{Q})}, \theta_m^{(\mathbb{Q})}, \nu_m^{(\mathbb{Q})}, \nu_v^{(\mathbb{Q})}, \mu_Y^{(\mathbb{Q})}, \sigma_Y^{(\mathbb{Q})}\}. \tag{21}$$

The remaining parameters are equal under both measures:

$$\Theta^{\mathbb{P},\mathbb{Q}} = \{\lambda_0^Y, \lambda_1^Y, \lambda_2^Y, \lambda_0^v, \lambda_1^v, \lambda_2^v, \lambda_1^m, \lambda_2^m, \sigma_m, \sigma_v, \rho_{Yv}, \rho_J\}. \tag{22}$$

The vector of all parameters is then $\Theta = \{\Theta^{\mathbb{P}}, \Theta^{\mathbb{Q}}, \Theta^{\mathbb{P},\mathbb{Q}}\}$.

5.1 Least-squares calibration - Methodology

In this approach, we calibrate our model to the cross section of S&P 500 and VIX options on some chosen dates. On each date, the output will be a set of values for the risk-neutral parameters $\Theta^{\mathbb{Q}}$ and $\Theta^{\mathbb{P},\mathbb{Q}}$. Calibration to one single day of options data does not allow us to estimate the parameters $\Theta^{\mathbb{P}}$ since options are priced under the pricing measure $\mathbb{Q}$ and no time series is used. This exercise is important because if the model is not able to reproduce well the implied volatility patterns of both markets together on a single date, then there is no point in estimating the model using a filter on a time series of options and indices.

We fix a date t. Let us consider $\{IV\_SPX_i^{Mkt}\}_{i\in I}$ the set of implied volatilities of options on the S&P 500 for the strikes $\{K_i\}$ and maturities $\{T_i\}$ available in our dataset[5] for this date. We use the superscript Mkt for 'Market' implied values. We denote by $\{IV\_VIX_j^{Mkt}\}_{j\in J}$ the set of VIX option implied volatilities on the same date t. $I = \{1, ...\#I\}$ is the set of integers indexing S&P 500 options available for this date and J the set indexing VIX options. To estimate parameter values, we minimize the distance between market and model implied volatilities (or option prices equivalently). We have chosen two distance criteria[6] that put different emphasis on S&P 500 and VIX options as well as on at-the-money (ATM) and out-of-the-money (OTM) options. We denote by $IV\_SPX_i^{Mod}$ the implied volatility of option with strike K_i and maturity T_i (respectively $IV\_VIX_j^{Mod}$ corresponding to the

[5]The dataset is described in the empirical analysis section 3 where we explain how implied volatilities have been calculated from S&P 500 and VIX options.

[6]We do not consider other popular choices of distances including absolute error of the logarithm of option prices, relative error of option prices (see Christoffersen and Jacobs (2004)) because we analyze the fits in section 6 depending on the implied volatilities and not option prices. Alternatively, we checked that using distances taking into account the bid-ask spread of IVs as in Cont and Kokholm (2011) does not significantly change the quality of fits.

notations above). The root mean squared error (RMSE) in implied volatilities on date t is defined as:

$$\begin{aligned} RMSE\_SPX(t) &:= \sqrt{\frac{1}{\#I}\sum_{i\in I}\left(IV\_SPX_i^{Mkt} - IV\_SPX_i^{Mod}\right)^2} \\ RMSE\_\mathrm{VIX}(t) &:= \sqrt{\frac{1}{\#J}\sum_{j\in J}\left(IV\_\mathrm{VIX}_j^{Mkt} - IV\_\mathrm{VIX}_j^{Mod}\right)^2} \\ RMSE(t) &:= \frac{1}{2}\left(RMSE\_SPX(t) + RMSE\_\mathrm{VIX}(t)\right) \end{aligned} \tag{23}$$

We furthermore consider the average relative error (ARE) in implied volatilities on date t:

$$\begin{aligned} ARE\_SPX(t) &:= \frac{1}{\#I}\sum_{i\in I}\frac{\left|IV\_SPX_i^{Mkt} - IV\_SPX_i^{Mod}\right|}{IV\_SPX_i^{Mkt}} \\ ARE\_\mathrm{VIX}(t) &:= \frac{1}{\#J}\sum_{j\in J}\frac{\left|IV\_\mathrm{VIX}_j^{Mkt} - IV\_\mathrm{VIX}_j^{Mod}\right|}{IV\_\mathrm{VIX}_j^{Mkt}} \\ ARE(t) &:= \frac{1}{2}\left(ARE\_SPX(t) + ARE\_\mathrm{VIX}(t)\right). \end{aligned} \tag{24}$$

Since the implied volatilities (IVs) are the highest for OTM SPX puts and OTM VIX calls, the RMSE puts more emphasis on fitting these options (which are the most liquid together with ATM options). On the other hand, it is also arguable that a 1% absolute error on an IV does not have the same importance if the market IV is 10% or 80%. The average relative error distance ARE takes this consideration into account by computing relative errors.

To cope with the ill-posedness of the calibration problem and the potential existence of multiple minima, we use two global optimizers namely the Covariance Matrix Adaptation Evolution Strategy (CMA-ES), introduced by Hansen and Ostermeier (1996), and the Differential Evolution (DE) algorithm introduced by Storn (1996).[7] They are evolutionary algorithms designed for high-dimensional non-linear non-convex optimization problems in a continuous domain. They is based on the principle of biological evolution, i.e., at every step new parameters are generated based on the optimal set of parameters up to that step, the objective function is evaluated for each of these new sets, and the new optimal parameter set becomes the one which yields the smallest value.

5.2 Particle filter

While the least-square calibration provides us with a static estimation of parameter values, it is more insightful to use the whole time series of option and index prices to learn about the dynamic properties of the unobservable processes (volatility, central tendency, jumps) and the risk premia associated to them. Sequential Monte-Carlo methods are ideal for this purpose as they allow to sequentially filter the trajectories of latent processes based on the information available. As they take as input a time series

[7]We are grateful to Jochen Krause for his implementation of various evolution optimizers including the CMA-ES and DE algorithms.

of observations, they furthermore allow to better identify parameters and therefore deliver more robust estimates. We discretize our continuous-time model under $\mathbb{P}$ over a grid of times $t \in \{0, T_1, ..., T_N < \infty\}$ to obtain the corresponding state-space problem:

$$\begin{aligned}\Delta Y_t = Y_{t+1} - Y_t = [-\lambda^Y(v_t, m_t)(\theta_Z^{(\mathbb{P})}(1,0,0) - 1) - \frac{1}{2}v_t + \gamma_t]\Delta t \\ + \sqrt{v_t}\Delta W_t^{Y(\mathbb{P})} + Z_t^{Y(\mathbb{P})}\Delta N_t^{Yv} \end{aligned} \tag{25}$$

$$\Delta v_t = v_{t+1} - v_t = \kappa_v^{(\mathbb{P})}\left(\frac{\kappa_v^{(\mathbb{Q})}}{\kappa_v^{(\mathbb{P})}}m_t - v_t\right)dt + \sigma_v\sqrt{v_t}\Delta W_t^{v(\mathbb{P})} + Z_t^{v(\mathbb{P})}\Delta N_t^{Yv} \tag{26}$$

$$\Delta m_t = m_{t+1} - m_t = \kappa_m^{(\mathbb{P})}(\theta_m^{(\mathbb{P})} - m_{t^-})dt + \sigma_m\sqrt{m_{t^-}}\Delta W_t^{m(\mathbb{P})} + Z_t^{m(\mathbb{P})}\Delta N_t^m. \tag{27}$$

The set of latent factors is: $L_t = \{v_t, m_t, \Delta N_t^{Yv}, \Delta N_t^m, Z_t^{Y(\mathbb{P})}, Z_t^{v(\mathbb{P})}, Z_t^{m(\mathbb{P})}\}$. Note that among these factors, only v_t and m_t are dependent on their past values, as the jump sizes are independent and identically distributed (i.i.d.) and so are the increments of Poisson processes conditionally on v_t and m_t. Equation (25) is the first measurement equation, the second is given by the VIX following equation (12). Other observable variables include the prices of options on the S&P 500 and VIX. We denote the set of observable variables used at time t by y_t. We assume that option prices are observed with an error, which might be due to different sources including the bid-ask spread, processing errors and timing (all options considered in one day are not traded at the same time) and misspecification error. The observation equations for options are given in equations (28) and (29):

$$\frac{O_{t,i}^{SPX,model}(Y_t, v_t, m_t, \Theta^{\mathbb{Q}}, \Theta^{\mathbb{P},\mathbb{Q}}) - O_{t,i}^{SPX,market}}{O_{t,i}^{SPX,market}} = \epsilon_{t,i}^{SPX,options} \tag{28}$$

$$\frac{C_{t,j}^{\text{VIX},model}(v_t, m_t, \Theta^{\mathbb{Q}}, \Theta^{\mathbb{P},\mathbb{Q}}) - C_{t,j}^{\text{VIX},market}}{C_{t,j}^{\text{VIX},market}} = \epsilon_{t,j}^{\text{VIX},options} \tag{29}$$

where $O_{t,i}^{SPX,market}$ corresponds to the price at time t of an option on the S&P 500 market indexed by $i \in I$. With a similar notation, $O_{t,i}^{SPX,model}(Y_t, v_t, m_t, \Theta^{\mathbb{Q}}, \Theta^{\mathbb{P},\mathbb{Q}})$ is the price given by the model with $\mathbb{Q}$ parameters $\{\Theta^{\mathbb{Q}}, \Theta^{\mathbb{P},\mathbb{Q}}\}$. $C_{t,j}^{\text{VIX},market}$ denotes the market price of the call option on the VIX indexed by $j \in J$. We assume the error terms to be normally distributed and heteroscedastic:

$$\epsilon_{t,i}^{SPX,options} \sim \mathcal{N}(0, \sigma^2_{\epsilon_{t,i}^{SPX}}), \tag{30}$$

and

$$\epsilon_{t,j}^{\text{VIX},options} \sim \mathcal{N}(\mu_{\epsilon_t^{\text{VIX}}}, \sigma^2_{\epsilon_{t,j}^{\text{VIX}}}), \tag{31}$$

where $\mu_{\epsilon_t^{VIX}}$ is proportional to the error which has been made on the estimation of the VIX level. Indeed, if the underlying's value is not accurately estimated, it introduces a systematic bias on the calculation of VIX option prices. We specify the variance of errors as follows:

$$\sigma^2_{\epsilon^{SPX}_{t,i}} = \exp\left(\phi_0 \text{bid-ask spread}_i + \phi_1 \left|\log\left(\frac{K_i}{F_t^{SPX}(T_i)}\right)\right| + \phi_2(T_i - t) + \phi_3\right) \tag{32}$$

$$\sigma^2_{\epsilon^{\text{VIX}}_{t,j}} = \exp\left(\psi_0 \text{bid-ask spread}_j + \psi_1 \left|\log\left(\frac{K_j}{F_t^{\text{VIX}}(T_i)}\right)\right| + \psi_2(T_j - t) + \psi_3\right). \tag{33}$$

with ϕ_i and ψ_i are in $\mathbb{R}$, $i \in \{0, ..., 3\}$.

We also assume that VIX levels are observed with error. Indeed, as they are in practice calculated using a finite number of options, a discretization bias is introduced. Furthermore, Jiang and Tian (2007) point to systematic biases in the VIX. We write this error as follows:

$$\text{VIX}_t^2 - (\alpha_{\text{VIX}^2} v_t + \beta_{\text{VIX}^2} m_t + \gamma_{\text{VIX}^2}) = \epsilon_t^{\text{VIX}}. \tag{34}$$

The error terms ϵ_t^{VIX} are assumed to follow a normal distribution with mean zero and variance $s > 0$.

The log-likelihood of a time-series of observations $y^n = (y_1, ..., y_n)$ with joint density p conditionally on a set of parameters Θ and a model specification $\mathcal{M}$ is equal to:

$$\log L(y^n|\Theta, \mathcal{M}) = \log p(y_1, ..., y_n|\Theta, \mathcal{M}) = \sum_{t=2}^{n} \log p(y_t|y^{t-1}, \Theta, \mathcal{M}) + \log p(y_1|\Theta, \mathcal{M}) \tag{35}$$

where

$$p(y_t|y^{t-1}, \Theta, \mathcal{M}) = \int p(y_t|L_t, \Theta, \mathcal{M}) p(L_t|y^{t-1}, \Theta, \mathcal{M}) dL_t. \tag{36}$$

Given an initial density $p(L_0|\Theta, \mathcal{M})$, the transition density of state variables $p(L_t|L_{t-1}|\Theta, \mathcal{M})$ and the likelihood function $p(y_t|L_t, \Theta, \mathcal{M})$, filtering methods allow to estimate the distribution $p(L_t|y^t, \Theta, \mathcal{M})$ of the current state at time t given the observations up-to-date. In particular, particle filters are perfectly adapted to our problem since they can handle nonlinear problems with non-Gaussian innovations. The key idea is to represent the required posterior density function of the latent variables $p(L_t|y^t, \Theta)$ by the empirical point-mass estimate $\hat{p}(L_t|y^t, \Theta, \mathcal{M})$:

$$\hat{p}(L_t|y^t, \Theta, \mathcal{M}) = \sum_{i=1}^{np} \pi_t^{(i)} \delta(L_t - L_t^{(i)}) \tag{37}$$

where $\pi_t^{(i)}$ denotes the normalized importance weight for particle i and $\delta(.)$ is the Dirac function.

To apply the particle filter, ones needs to be able to simulate at every time t a number np of particles $L_t^{(i)}, i = 1..np$ from $p(L_t|y^{t-1}, \Theta, \mathcal{M})$ and to be able to calculate $p(y_t|L_t^{(i)}, \Theta, \mathcal{M})$. Based on the simulated particles, $p(y_t|y^{t-1}, \Theta, \mathcal{M})$ is approximated as:

$$p(y_t|y^{t-1},\Theta,\mathcal{M}) \approx \frac{1}{np}\sum_{i=1}^{np} p(y_t|L_t^{(i)},\Theta,\mathcal{M}). \tag{38}$$

Multiple versions of the particle filter exist. We use the Auxiliary Particle Filter (APF) proposed by Pitt and Shephard (1999) and extend the approach described in Johannes, Polson, and Stroud (2009). The main advantage of the APF compared to more basic particle filters such as the Sampling Importance Resampling (SIR) is that it is able to handle jumps while the SIR filter faces the problem of sample impoverishment leading to particle degeneracy. Both filters are described in Johannes, Polson, and Stroud (2009) for filtering latent factors from returns in a Heston model with jumps. The authors also use it with option prices but do not provide details on the adjustments made. Our extension of the filter makes it possible to extract the most probable paths of both factors v and m as well as the jump components from the set of observable variables y^n. To incorporate the information contained in the S&P 500 and VIX levels, we calculate at every time t the joint probability of having 0 or 1 jump in every process given the new observations. As the jump sizes of the forward returns are normally distributed but those in the variance processes are exponentially distributed, the likelihood of the new observations given a combination of jumps involves a sum of normal and exponential random variables. To approximate the joint probability of jumps and keep tractability, we represent the exponential law by a finite set of values taken at some chosen quantiles.[8] To overcome the usual difficulties of particle filters to recognize jumps, we furthermore systematically simulate jumps in one tenth of the particles.[9]

The detailed filtering procedure is described in Appendix 5.

6 Daily calibration results

The first step in evaluating the performance of the model (1)-(3) is to calibrate it to one day of options data to make sure that the model is flexible enough to simultaneously price options on both markets.

We follow the method outlined in section 5.1. We have chosen some dates on which we calibrate the model (1)-(3) to the cross section of S&P 500 and VIX implied volatilities. We only report the results on four dates as they are representative of the whole sample. We consider two days where the market was facing great uncertainty about the future, October 22 2008 (one month after the bankruptcy of Lehman Brothers) and May 05 2010, at the beginning of the European sovereign debt crisis. On these dates, the markets were under stress and S&P 500 implied volatilities had a very strong negative skew and levels above 100% for short term options. The other two days we report are rather calm compared to those: July 11 2007 and June 10 2009.

Gatheral (2008) has shown that the Heston model is incapable of reproducing the positive skew in VIX implied volatilities (IVs) as displayed in Figure 4 and Sepp (2008a,b) added that introducing positive jumps in the volatility dynamics of the Heston model allows the model to have a positive skew

[8] Robustness tests were performed on simulated data to check that the choice of quantiles was appropriate.
[9] We tested the accuracy of the filter using different proportions of particles with systematic jumps and found that 1/10 was a reasonable choice.

in VIX IVs. As a consequence, departing from the usual literature on S&P 500 option pricing, the simplest model we consider is the Heston model with jumps in returns and volatility. It corresponds to the model (1)-(2) where the central tendency m is constant. We will denote this model by SVJ (Stochastic Volatility with Jumps). The most flexible model we consider, with 2 factors to represent the volatility component namely the variance and a stochastic central tendency, is referred to as SVJ2.

[Insert Figure 4 here]

We report in Table 3 the results for the RMSE calibration with respect to implied volatilities (23).[10] We emphasize that for each model and on each day, we have minimized the total RMSE from the VIX and the S&P 500 market together. We report the resulting $RMSE\_SPX$ and $RMSE\_VIX$ since these are indicative of the quality of the fit on each market. Irrespective of the day, we observe that the SVJ and SVJ2 models perform comparably on the S&P 500 options market, both fitting very well with an average RMSE of around 1.5% across the dates we have calibrated to. In contrast, we see that there are dates when the SVJ model struggles to fit the VIX IVs in addition to the S&P 500 IVs whereas the SVJ2 model satisfactorily fits both. This is observed for instance on July 11 2007 and on May 05 2010. We can see the comparative fits for the VIX options market in Figure 5.[11] Panels A to D correspond to the SVJ2 model's fit and panels E to H to the SVJ model's fit. On this date, it seems that for short maturities, the SVJ model has difficulties to reproduce the strong positive skewness of VIX IVs (which was already the case for the Heston model). This shortcoming of the SVJ is not often noticeable and we therefore do not make it a general statement. Indeed, on the other two dates we report, the fits of the SVJ and SVJ2 are comparable on the VIX market.

[Insert Table 3 here]

[Insert Figure 5 here]

Daily calibration is essentially a multiple curve fitting exercise, where we check whether the models can fit the risk-neutral distributions inferred by option prices at different maturity. A more thorough analysis is needed to conclude at this point that the SVJ2 is much better than the SVJ model to price S&P 500 and VIX options together. Indeed the SVJ2 has more parameters and is therefore bound to fit better. Furthermore, some of the parameters we get from daily calibrations can vary a lot from one day to the next.[12] At this point, it is therefore not possible to know whether the dynamics of the model (1)-(3) can reproduce the time evolution of these smiles. This is what we will focus in the next section.

[10] The results are qualitatively similar when minimizing over the distances (23) and (24), we therefore only report one result.

[11] The SVJ and SVJ2 both match the S&P 500 options market prices almost perfectly so we do not show them.

[12] As explained in Broadie, Chernov, and Johannes (2007) and Lindström, Ströjby, Brodén, Wiktorsson, and Holst (2008), the parameters obtained when calibrating to daily options prices are not stable over time. To better understand how the model performs over time, it is important to estimate the model on a time series of options data.

7 Time-series estimation results using a Particle Filter

The second step in evaluating the performance of the model is to estimate it using the time series of S&P 500 and VIX indices together with S&P 500 and VIX options. This is achieved using the particle filter described in Section 5.2. We report the results for different sub-models of (1)-(3) and analyze the gain of information and robustness we have from adding the data from the VIX market to the dataset. The sub-models considered of the SVJ2 model are the SVJ model and the 2-factor continuous model, which has a stochastic central tendency but no jumps. We refer to the latter model as SV2.

7.1 Specific data treatment for the Particle filter

The dataset contains a large amount of ATM options compared to OTM and deep OTM options. This implies that if we use the filter on this dataset and the model is not able to fit all options, the fitting of ATM options will be its priority rather than deep OTM options. Given the formula in Breeden and Litzenberger (1978), this results in fitting the body of the S&P 500 returns distribution rather than the tails which is not what we want: We need information about extreme events contained in the data to be incorporated into the models. For this reason, we have decided to interpolate the S&P 500 slices and to re-sample option prices from our parametric fit uniformly across moneynesses.[13] Other advantages of our interpolating method is that the resulting data is arbitrage free,[14] we have fewer points for each slice (still representing accurately the information of each slice), thus reducing computational complexity.

We explain in detail in Appendix 4 how we have used the efficient mixture of log-normals approach of Rebonato and Cardoso (2004) to have a parametric fit for each S&P 500 implied volatility slice. The root mean squared error of the S&P 500 implied volatilities parametric fits are on average around 0.25% and we therefore do not loose information especially given the market bid-ask spread. Finally, given the parametric fit for a given slice, we sample a fix number (we have chosen 15) of option prices uniformly distributed on the moneyness axis.

We do not perform any interpolation for the VIX options dataset as most VIX options are OTM and therefore contain information about the tails of the VIX distribution (i.e., variance and central tendency processes).

We divide the data into four different datasets:

Dataset 1: S&P 500 returns and VIX index levels,

Dataset 2: S&P 500 returns, VIX index levels and S&P 500 options,

[13] It is common to interpolate data, see, e.g., Broadie, Chernov, and Johannes (2007). This eliminates arbitrage opportunities in the data and removes the accumulation of options in the ATM region.

[14] Since we have considered mid-prices and because of synchronization issues between the underlying and the options, implied volatility slices are a priori not guaranteed to be arbitrage free.

Dataset 3: S&P 500 returns, VIX index levels and VIX options,

Dataset 4: S&P 500 returns, VIX index levels, S&P 500 options and VIX options.

Using these different datasets allows us to make inference on what information they contain, and whether they are consistent with one another. Our filter uses daily time steps. It incorporates information on the underlying indices on a daily basis. As the database comprises a large amount of options, it is unfeasible to calculate option prices every day for every particle, we follow Pan (2002) and Johannes, Polson, and Stroud (2009) among others and use weekly (Wednesday) option data. Furthermore, this eliminates beginning-of-week and end-of-week effects.

The time-series of observations is decomposed into two periods, the first one from March 1st, 2006 to October 10, 2008 (before the peak of crisis and the VIX index increased to its highest point). This is a rather calm period,[15] that we will use as in-sample estimation period. Our out-of-sample period starts on October 11, 2008 and ends on October 29, 2010. This period includes very high levels of volatility (i.e., implied volatilities from S&P 500 and VIX options as well as VIX index values). The last column of Table 4 reports the amount of options within each moneyness and maturity range in both periods.

[Insert Table 4 here]

7.2 Filtered trajectories

Figure 7 displays the filtered trajectory of the volatility process when estimating the SVJ2 model over *Dataset 4*. Although the volatility trajectories are consistent across models, we notice that the volatility is slightly more variable when the calibration is done with models with a stochastic central tendency. Indeed, the Feller condition is imposed on the SVJ model, which restricts the amplitude of volatility movements. This condition is relaxed for 2-factor models as the long-term mean of the variance is varying, which allows the volatility process to have a larger amplitude. In particular, for all datasets the estimated parameter σ_v, which controls the volatility of volatility, is significantly smaller for the SVJ model than for the SV2 and SVJ2 models, see Table 5. Figure 8 represents the difference between the filtered volatility processes using *Dataset 4* and the the other datasets (*Datasets 1* to *3*). Until the peak in the VIX toward the end of the in-sample period, this difference is very small (lower than 2%). In this period, the filtered volatility using *Dataset 1* remains slightly lower than the volatility filtered using the other datasets, therefore options contribute to increase the filtered volatility, in line with the literature. During the peak of the VIX end of 2008, the volatility filtered using *Dataset 4* is on average close to the one filtered using *Dataset 3* but up to 23% lower than the volatility filtered using *Datasets 1* and *2*. Therefore adding S&P 500 options to *Dataset 1* does not provide much new information on the behaviour of the variance. Reversely, adding VIX options brings information on the volatility in times of market turmoil which is not spanned by the underlying levels

[15] We have decided to include the beginning of October 2008 so that the in-sample period actually includes several dates with extreme events.

and the S&P 500 options. In the out-of-sample period, the difference between the trajectories remains within ±3% except during the short peak of variance in 2010, where there is another peak in volatility but of smaller magnitude. Figure 9 shows the jump sizes when then the filtered jump probability is larger than 50%. Small jumps (around 2%) are filtered in the variance process beginning of 2007, and larger jumps (above 10%) are filtered towards the end of 2008 as the VIX peaks, when VIX options are not part of the dataset. Using *Dataset 3* the algorithm does not filter any jump in v, but finds these jumps more likely to occur in the trajectory of m, see Figure 10.

[Insert Figures 7, 8, 9 and 10 here]

[Insert Table 5 here]

The process m exhibits comparable shapes across estimation datasets and 2-factor models, see Figure 11. However, we notice that when options are part of the estimation dataset, the magnitude of m increases. Using *Dataset 1*, m reaches a maximum of about 35% in 2009, but it goes up to 45% when S&P 500 or VIX options are added to the dataset. The trajectories filtered by the SV2 and SVJ2 models almost overlap before the crisis, however the increase occurs earlier with the SVJ2 model than with the SV2 model, suggesting that including jumps in the model allows to be more reactive and adjust the central tendency faster. The process m is more stable and less erratic than the variance process, giving evidence that it captures long-term trends. Its volatility parameter is below 30% in all estimations. Its speed of mean-reversion is more than three times lower than the one of v under $\mathbb{P}$ and more than six times lower under $\mathbb{Q}$. The process is therefore more persistent. While the variance process increases dramatically during the crisis but then returns to a level which is comparable to the one before the crisis, the central tendency also increases during the crisis but then goes down to a level which is higher than before the crisis (between 25 and 30%). Therefore v captures the punctual movements of the variance while m, as the stochastic long-term mean of v, reflects the longer-term expectations of investors regarding the variance and can therefore be seen as an indicator of market turmoil.

[Insert Figure 11 here]

7.3 Are jumps and/or a stochastic central tendency needed?

In this part we analyze whether jumps and a stochastic central tendency are needed to reproduce the features of VIX levels, S&P 500 option prices and finally VIX option prices.

Let us first investigate whether these features are needed to provide an accurate fit of the VIX index. Given the filtered trajectories for the processes v and m inferred by *Dataset 1* we calculate the corresponding model-implied values of the VIX index using the optimal parameters. As illustrated by Figure 12, the model accurately reproduces the time-series of VIX index values. Table 6 reports the corresponding Mean Errors (ME) and Root Mean Square Errors (RMSE) and shows that they are

comparable across models. This observation is consistent across datasets. At first sight, jumps and a stochastic central tendency therefore seem superfluous to reproduce the trajectory of the VIX level.

[Insert Figure 12 here]

[Insert Table 6 here]

To statistically challenge this claim, we run likelihood tests. Table 5 reports the log-likelihood values as well as the values of the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) for *Datasets 1* to *4*. Both criteria are slightly in favor of the SVJ2 model and therefore support the use of jumps and of a stochastic central tendency.

We further challenge the in- and out-of-sample performance of the SVJ2 model by running various Diebold-Mariano (DM) tests. For the three models considered, we consider two loss functions namely the absolute and quadratic pricing errors, respectively defined as $L(e_t) = |e_t|$ and $L(e_t) = e_t^2$, where e_t refers to the difference at time t between the model-forecast of the VIX and the true value of the index. We denote the loss differential at time t between the SVJ or the SV2 model and the SVJ2 model by $d^{SVJ/SV2,SVJ2} = L(e_t^{SVJ/SV2}) - L(e_t^{SVJ2})$. If the two models considered have comparable pricing errors, then $\mathbb{E}^{\mathbb{P}}[d^{SVJ/SV2,SVJ2}] = 0$. A positive value of the expectation means that the SVJ2 model outperforms the sub-model considered. Table 7 reports the results when the calibration is done using *Dataset 1* and shows that the test values are very close to zero with the quadratic loss function and negative with the absolute loss function, suggesting that the SVJ and SV2 models should be preferred to the SVJ2 model when calibrating them to underlying levels only. When including options in the estimation dataset, we obtain test values which are very close to zero (within ± 0.1 bounds). Therefore we conclude that the three models considered perform comparably at reproducing the trajectories of VIX levels, in- and out-of-sample.

[Insert Table 7 here]

The calibration to S&P 500 options (*Dataset 2*) highlights the superiority of the SVJ2 and SV2 models over the SVJ model. Computationally, the SV2 model is faster to calibrate as the detection and estimation of jumps involves rare events and is challenging for the particle filter. But more importantly, the SVJ model exhibits Root Mean Square Errors (RMSEs) and Root Mean Square Relative Errors (RMSREs) when pricing S&P 500 options which are for most option categories higher than those of the SV2 model. In particular, the SVJ model does not represent well the deep OTM and long-maturity options, see Tables 4 and 8. The corresponding RMSREs are almost twice those of the 2-factor volatility models for in-sample deep OTM calls and about twice for in-sample long-maturity options. In the out-of-sample period, the poorer performance of the SVJ model extends to medium-maturity options and OTM calls. Therefore a stochastic central tendency allows to better price long-term and deep OTM S&P 500 options. This supports the hypothesis that the process m captures the long-term trends of volatility, and therefore enables to better reproduce the term structure of S&P 500 option prices.

[Insert Table 8 here]

Using the Diebold-Mariano test, we test whether the SVJ2 model's pricing performance of SPX options is significantly better than that of its sub-models. For this purpose, we consider two loss functions, the Mean Square Error (MSE) of option prices and the Mean Square Relative Error (MSRE). The resulting test values in Table 10 associated to S&P 500 options (*Dataset 2*) are strongly positive in-sample indicating that the SVJ2 model significantly outperforms the SVJ and SV2 models. Out-of-sample, we note that the SV2 model performs comparably to the SVJ2 model when using *Dataset 4* as estimation dataset.

The calibration to VIX options (*Dataset 3*) also favors the SVJ2 model, which yields better RMSEs and RMSREs of VIX option prices than the SVJ and SV2 models except for deep OTM options. The comparison of the SV2 and SVJ models shows that the SVJ model better prices these deep OTM calls while the SV2 model is more appropriate for other moneyness levels. The SVJ model can therefore better represent the tail of the volatility distribution. Consistently with the results we obtained when estimating models to *Dataset 2*, the SV2 model outperforms the SVJ model in pricing the medium-maturity VIX options (which are the longest maturity VIX options).

Again we run Diebold-Mariano tests to statistically evaluate whether the SVJ2 model improves upon its sub-models at pricing VIX options. The MSE test values are all very close to zero. But the MSRE test values are consistently larger than zero in-sample, suggesting that on average the SV2 and the SVJ models are slightly outperformed by the SVJ2 model, see Table 9. Out-of-sample, as before the SV2 model appears to be slightly preferrable to the SVJ2 model.

[Insert Table 9 here]

When including options on both markets in the estimation dataset (*Dataset 4*), the SVJ2 model yields RMSEs and RMSREs which are smaller than the SVJ and the SV2 models in-sample in pricing most S&P 500 and VIX option categories, see Tables 4 and 8. We notice that the SVJ model performs particularly bad in pricing the deep OTM options on the S&P 500, consistently with the results obtained when calibrating the models to *Dataset 2*. The Diebold-Mariano tests confirm that the SVJ2 model provides significantly better in-sample MSREs for S&P 500 options than the two other models, see Table 10. We however note that in the out-of-sample period, the SV2 model yields smaller errors than the SVJ2 model for VIX options, which might be due to a problem of identification of the jump terms.

Therefore we conclude that a stochastic central tendency is necessary to accurately price long-term options and represent the tails of the returns' distribution (OTM calls on the S&P 500). On the other side, jumps add value to represent the right tail of the variance distribution (OTM calls on the VIX). Therefore the full flexibility of the SVJ2 model is needed to represent the underlying returns in a way that is consistent with S&P 500, VIX levels and their derivatives prices. This conclusion is however mitigated by the difficulty to identify jump terms.

Furthermore we observe that the SVJ2 model encounters problems during the crisis and does not

well represent volatility smiles. In particular, OTM puts on the S&P 500 tend to be underpriced and OTM calls are generally overpriced, i.e., the smile of volatility does not exhibit enough skewness. This phenomenon affects short-maturity options in particular. Figure 13 compares the moments of S&P 500 returns as implied by market and model option prices, when the models are calibrated to *Dataset 4* (all indices and options). While the skewness of the returns is well represented at the beginning of the in-sample period, it is underestimated from late 2007 until the end of our sample. In the out-of-sample period this phenomenon becomes much more apparent, and all three models yield an implied skewness which is about half the one implied by the market. Similarly, the kurtosis is only slightly underestimated at the beginning of the time-series, but in the out-of-sample period the model kurtosis it is about half the market implied kurtosis. We also note that there is no improvement in the representation of SPX implied moments when adding the options on the VIX market to the estimation dataset.

[Insert Figure 13 here]

Figure 14 displays the implied moments of VIX returns and compares the market values to the model values of these moments. From the end of 2007 we notice that the SVJ model yields very variable moments, which are very far from those implied by the market. In particular, the sign of the skewness is very often not correctand the kurtosis reaches very large values exceeding 20. The SVJ2 and SV2 models perform significantly better, however their ability to represent the skewness and kurtosis drastically deteriorates at the very end of the in-sample period. In times of market turmoil where the VIX exhibits peaks, they even produce negative skewness for VIX options, which is at odds with empirical evidence, even though the VIX IVs tend to become flat during the crisis (see Figure 2, VIX IVs on 27/10/2008). The VIX negative skewness of the SVJ2 model is interestingly not visible in the static calibration presented in Section 6 and might be an issue that appears when trying to consistently estimate the model over a long period of time. However, we keep in mind that this result might be driven to some extent by the small sample of VIX options at our disposal, when ideally the non-parametric formula of Bakshi, Kapadia, and Madan (2003) that we employ ideally requires to use a continuum of options. Finally, in January 2009 the model implied kurtosis is more than three times equal to its market counterpart.

7.4 Information contained in the different data sources

In this part we address the two following questions: first, what information do levels contain on options and second, which information do S&P 500 and VIX options share?

Since the values of the VIX index are calculated using a portfolio of S&P 500 options, it is tempting to see the VIX as a summary of the information contained in S&P 500 options. Following such reasoning, it is interesting to see to which extent the model estimated using *Dataset 1* is able to reproduce options' prices. We obtain RMSEs and RMSREs that are respectively more than twice those obtained using *Dataset 4* for both S&P 500 and VIX options. Therefore our results clearly indicate that calibrating the model to the underlying index values is not sufficient to reproduce options' prices in either market.

Let us now examine the information content of VIX and S&P 500 options. As emphasized in Section 7.2, the trajectories of the processes v and m obtained using *Datasets 2* and *3* are very similar. Therefore S&P 500 and VIX options contain consistent information on the evolution and behavior of these processes. We however notice that jumps are filtered in m only when S&P 500 options are not part of the estimation dataset, which suggests that VIX and S&P 500 options provide conflicting information on the jump component of m.

Furthermore, it is interesting to see that the RMSEs and RMSREs of VIX options using *Datasets 2* are about twice as large as those using *Dataset 3*, see Tables 4 and 8. This consideration holds in- and out-of-sample. Thus we conclude that S&P 500 options do not span the information contained in VIX options.

Conversely, the RMSEs and RMSREs of S&P 500 options are overall much lower (reduced by a factor 2 and 4 approximately) using *Dataset 2* than using *Dataset 3*. This is in particular due to deep OTM calls that are not well reproduced at all using *Dataset 3*, which indicates that VIX options contain less information on the tail of the returns' distribution than S&P 500 options. Concerning deep OTM puts on the S&P 500, it is striking to see that the estimation using *Dataset 3* outperforms the one using *Dataset 2*, which indicates that VIX options provide valuable information on the left tail of the returns' distribution.

We however notice that there is a significant loss of quality in the fitting of VIX options when estimating the SVJ2 model to all data sources, which indicates that the SVJ2 model is not flexible enough to reconcile the information contained in all data sources in a completely consistent way. This contrasts with the results obtained in the static calibrations performed in section 6.

7.5 Risk premia

Let us analyze the signs of the risk premia defined in Section 4.2. The equity risk premium coefficient η_Y is found to be positive throughout the models and datasets considered, which is in line with a positive diffusive equity risk premium. When options are part of the estimation dataset, it is consistently between 0.6 and 0.85. The mean price jump risk premium is found to be slightly negative when all data sources are reconciled, i.e., the mean jump size of returns is slightly les negative under $\mathbb{P}$ than under $\mathbb{Q}$, which is also what Pan (2002) finds. The volatility of price jump risk premium given by the full model is around 10%, indeed the volatility of jump sizes under $\mathbb{Q}$ is around 10-15% while it is around 2-3% under $\mathbb{P}$. The diffusive part of the volatility risk premium is found to be negative, its amplitude however depends on the model used. In particular, it has much smaller magnitude with the SVJ2 model than with the SVJ and SV2 models. Our results on the mean volatility jump risk premium are mitigated. It is indeed positive when using *Dataset 3* but negative when using *Dataset 4*. The average jump size under $\mathbb{P}$ increases when adding S%P 500 options while under $\mathbb{Q}$ the reverse occurs. However, the fact that no jump has been filtered using *Dataset 3* suggests that mean volatility jump risk premium is negative. Finally, the diffusive volatility risk premium in m is also found to be negative, i.e., the process means-reverts quicker under $\mathbb{P}$ than under $\mathbb{Q}$. We did not obtain conclusive

results on the central tendency jump risk premium.

We now examine integrated risk premia and their term structure as implied by the different models, calibrated over the four datasets previously considered.

The annualized integrated equity risk premium ERP_t at time t is defined as:

$$ERP_t = \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(\int_t^T dY_s\right) - \mathbb{E}_t^{\mathbb{Q}}\left(\int_t^T dY_s\right)\right].$$

The annualized integrated variance risk premium VRP_t can be decomposed into a continuous and a jump part as follows:

$$VRP_t^c = \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(\int_t^T v_s ds\right) - \mathbb{E}_t^{\mathbb{Q}}\left(\int_t^T v_s ds\right)\right],$$

$$VRP_t^j = \frac{1}{T-t}\left[\mathbb{E}_t^{\mathbb{P}}\left(\sum_s (\Delta Y_s)^2\right) - \mathbb{E}_t^{\mathbb{Q}}\left(\sum_s (\Delta Y_s)^2\right)\right],$$

$$VRP_t = VRP_t^c + VRP_t^j.$$

For a detailed discussion on risk premia we refer to Bollerslev and Todorov (2011). In our setup all risk premia are available in closed-form.

We obtain integrated equity risk premia that are positive and strongly increasing during the crisis in 2008, then coming back to a low level and suddenly increasing again end of 2010, following the VIX movements, see Figure 15. During these periods of market turmoil, shorter-maturity risk premia are larger than longer-maturity premia which is in line with Aït-Sahalia, Karaman, and Mancini (2012). Our results are consistent in sign and magnitude across estimation datasets. Furthermore we find that the equity risk premium is primarily determined by its continuous part. Indeed, the jump part of the daily risk premium is usually smaller than 0.1 while the continuous part reaches about 0.7.

Finally, we obtain integrated variance risk premia which are slightly negative, which a magnitude which increases in 2008. Consistently with the literature shorter-maturity risk premia are close to zero than longer-maturity premia. The one year risk premium reaches -0.05 during the volatility peak and is mainly determined by its jump component when using the SVJ and SVJ2 models.

[Insert Figure 15 here]

8 Conclusion

In this paper we estimate a general affine model with jumps using a time series of S&P 500 and VIX levels as well as option prices on both indices. To extract as much information about extreme events as possible, we use S&P 500 and VIX options with a unique wide range of moneynesses. We depart from most of the literature and estimate the historical $\mathbb{P}$-parameters and the risk-neutral $\mathbb{Q}$-parameters jointly, in a single step. This estimation puts less restrictions on the parameters and allows us to obtain results on risk premia which are purely data-driven. We show that although the standard SVJ model performs well at representing the smiles of volatility for both markets on a given date, its dynamics is not sufficiently flexible to accommodate for the dynamical properties embedded in the time series of option prices. We argue that a model with a stochastic central tendency and jumps in the returns and in each volatility factor brings significant improvements and allows to reach smaller pricing errors, both in- out-of-sample. We analyze the filtered trajectories of the latent factors using different estimation datasets and sub-models. We show that the variance process exhibits large and fast variations capturing the short-term movements of the volatility while the stochastic central tendency exhibits more persistence and hence reflects long-term expectations of investors. We provide an extensive analysis of which features of the SVJ2 are needed to represent different datasets. In particular, likelihood criteria as well as statistical tests conclude that the whole flexibility of the model is needed to jointly represent the index levels and the derivatives' prices on both markets. Indeed, adding a stochastic central tendency helps to better represent the tails of the returns' distribution as well as the term structure of S&P 500 and VIX option prices. Furthermore, jumps enable to better reflect the tail of the variance distribution. We highlight the limitations of the models considered, in particular we show that they are not able to fully reproduce the skewness and kurtosis of the underlying instruments in times of market turmoil. We investigate the information contained in the underlying levels and in the options on both markets, and find on the one side, that the VIX index does not provide an accurate representation of the information contained in S&P 500 options, and on the other side that the information contained in the S&P 500 derivatives does not span the information contained in the VIX derivatives and vice-versa. It is therefore crucial to include underlyings as well as derivatives on both markets to estimate a model and account for the cross section of instruments. We finally provide a discussion on the risk premia embedded in the model. We find that all the datasets considered provide consistent information on the equity risk premium and that it is mainly determined by its continuous component. We obtain variance risk premia which are slightly negative and on the contrary mainly affected by their jump part.

Appendix

1 Affine dependence of the VIX^2 on v and m

We provide the expression of the coefficients α_{VIX^2}, β_{VIX^2} and γ_{VIX^2} in Proposition 4.1. We can write

$$\begin{aligned}
\alpha_{\text{VIX}^2} &= \left(1 + 2\lambda_1^Y C\right) A \\
\beta_{\text{VIX}^2} &= \left(1 + 2\lambda_1^Y C\right) B + \left(2\lambda_2^Y C\right) \hat{A} \\
\gamma_{\text{VIX}^2} &= 2\lambda_0^Y C + \left(1 + 2\lambda_1^Y C\right) G + \left(2\lambda_2^Y C\right) \hat{B}.
\end{aligned}$$

where $C := \left(\theta_Z(1,0,0) - \frac{\partial \theta_Z}{\partial \phi_Y}(0,0,0) - 1\right)$ and coefficient $A, B, \hat{A}, \hat{B}, G$ are defined in Table 11. Furthermore, $A, B, \hat{A}, \hat{B}, G$ are functions of:

$$\begin{aligned}
a_m &:= \left(\frac{\partial \theta_Z}{\partial \phi_m}(0,0,0)\lambda_1^m - \kappa_m\right), \\
c_m &:= \left(\kappa_m \theta_m + \frac{\partial \theta_Z}{\partial \phi_m}(0,0,0)\lambda_0^m\right), \\
b_m &:= -\frac{c_m}{a_m}\ , \quad \text{when } a_m \neq 0, \\
a_v &:= \left(\frac{\partial \theta_Z}{\partial \phi_v}(0,0,0)\lambda_1^v - \kappa_v\right), \\
b_v &:= b_m \left(\kappa_v + \frac{\partial \theta_Z}{\partial \phi_v}(0,0,0)\lambda_2^v\right) + \frac{\partial \theta_Z}{\partial \phi_v}(0,0,0)\lambda_0^v\ , \quad \text{when } a_m \neq 0, \\
h_v &:= \left(\kappa_v + \frac{\partial \theta_Z}{\partial \phi_v}(0,0,0)\lambda_2^v\right).
\end{aligned}$$

These expressions remain valid with different specifications for the jump structure (dependent jumps, same Poisson processes, etc.) in the model (1) - (3).

2 Characteristic functions

The characteristic function of the processes Y, v and m defined in the model (1) - (3) are exponential affine as stated in Proposition 4.2.

$$\begin{aligned}
\Psi_{VIX_T^2}(t,v,m;\omega) &= \mathbb{E}_t\left[e^{\omega VIX_T^2}\right] = e^{\alpha(T-t)+\beta(T-t)v+\gamma(T-t)m}, \\
\Psi_{Y_T}(t,v,m;\omega) &= \mathbb{E}_t\left[e^{\omega Y_T}\right] = e^{\alpha_Y(T-t)+\beta_Y(T-t)y+\gamma_Y(T-t)v+\delta_Y(T-t)m},
\end{aligned}$$

where $\omega \in \mathbb{C}$ is in each case chosen so that the integral converges.

The coefficients entering the definition of Ψ_v satisfy the following ODEs:[16]

$$
\begin{aligned}
&-\alpha'(T-t)+\gamma(T-t)\kappa_m\theta_m+\lambda_0^v\left(\theta_Z(0,\beta(T-t),0)-1\right)+\lambda_0^m\left(\theta_Z(0,0,\gamma(T-t))-1\right)=0\\
&-\beta'(T-t)-\beta(T-t)\kappa_v+\frac{1}{2}\sigma_v^2\beta^2(T-t)+\lambda_1^v\left(\theta_Z(0,\beta(T-t),0)-1\right)=0\\
&-\gamma'(T-t)-\gamma(T-t)\kappa_m+\frac{1}{2}\sigma_m^2\gamma^2(T-t)+\kappa_v\beta(T-t)+\lambda_2^v\left(\theta_Z(0,\beta(T-t),0)-1\right)+\\
&\qquad\lambda_1^m\left(\theta_Z(0,0,\gamma(T-t))-1\right)=0
\end{aligned}
$$

$\forall t\in(0,T]$ with boundary conditions $\alpha(0)=0$, $\beta(0)=\omega_1$ and $\gamma(0)=\omega_2$, where $\omega_1:=\omega\alpha_{VIX^2}$ and $\omega_2:=\omega\beta_{VIX^2}$ (the coefficients α_{VIX^2} and β_{VIX^2} are defined in the Appendix 1).

The coefficients entering the definition of Ψ_Y satisfy different ODEs depending on the jump structure that we impose. When the jump processes entering the variance and in the returns are independent, then the ODEs are given by:

$$
\begin{aligned}
&-\alpha_Y'(T-t)+\beta_Y(T-t)(-\lambda_0^Y(\theta_Z(1,0,0)-1))+\delta_Y(T-t)\kappa_m\theta_m\\
&\qquad+\lambda_0^Y[\theta_Z(\beta_Y(T-t),0,0)-1]+[\theta_Z(0,\gamma_Y(T-t),0)-1]\lambda_0^v+\lambda_0^m[\theta_Z(0,0,\delta_Y(T-t))-1]=0\\
&-\beta_Y'(T-t)=0\\
&-\gamma_Y'(T-t)-\beta_Y(T-t)\lambda_1^Y(\theta_Z(1,0,0)-1)-\frac{1}{2}\beta_Y(T-t)-\gamma_Y(T-t)\kappa_v+\frac{1}{2}\beta_Y(T-t)^2\\
&\qquad+\frac{1}{2}\gamma_Y(T-t)^2\sigma_v^2+\beta_Y(T-t)\gamma_Y(T-t)\sigma_v\rho_{Y,v}+\lambda_1^Y[\theta_Z(\beta_Y(T-t),0,0)-1]+\lambda_1^v[\theta_Z(0,\gamma_Y(T-t),0)-1]=\\
&-\delta_Y'(T-t)-\beta_Y(T-t)\lambda_2^Y(\theta_Z(1,0,0)-1)+\gamma_Y(T-t)\kappa_v-\delta_Y(T-t)\kappa_m+\frac{1}{2}\delta_Y(T-t)^2\sigma_m^2\\
&\qquad+\lambda_2^Y[\theta_Z(\beta_Y(T-t),0,0)-1]+\lambda_2^v[\theta_Z(0,\gamma_Y(T-t),0)-1]+\lambda_1^m[\theta_Z(0,0,\delta_Y(T-t))-1]=0
\end{aligned}
$$

$\forall t\in(0,T]$ with boundary conditions $\alpha_Y(0)=0$, $\beta_Y(0)=\omega$, $\gamma_Y(0)=0$ and $\delta_Y(0)=0$.

When we assume that the Poisson process driving the jumps is the same in the variance process and in the returns process, then the ODEs are given by:

$$
\begin{aligned}
&-\alpha_Y'(T-t)+\beta_Y(T-t)(-\lambda_0^Y(\theta_Z(1,0,0)-1))+\delta_Y(T-t)\kappa_m\theta_m\\
&\qquad+\lambda_0^Y[\theta_Z(\beta_Y(T-t),\gamma_Y(T-t),0)-1]+\lambda_0^m[\theta_Z(0,0,\delta_Y(T-t))-1]=0\\
&-\beta_Y'(T-t)=0\\
&-\gamma_Y'(T-t)-\beta_Y(T-t)\lambda_1^Y(\theta_Z(1,0,0)-1)-\frac{1}{2}\beta_Y(T-t)-\gamma_Y(T-t)\kappa_v+\frac{1}{2}\beta_Y(T-t)^2\\
&\qquad+\frac{1}{2}\gamma_Y(T-t)^2\sigma_v^2+\beta_Y(T-t)\gamma_Y(T-t)\sigma_v\rho_{Y,v}+\lambda_1^Y[\theta_Z(\beta_Y(T-t),\gamma_Y(T-t),0)-1]=0\\
&-\delta_Y'(T-t)-\beta_Y(T-t)\lambda_2^Y(\theta_Z(1,0,0)-1)+\gamma_Y(T-t)\kappa_v-\delta_Y(T-t)\kappa_m+\frac{1}{2}\delta_Y(T-t)^2\sigma_m^2\\
&\qquad+\lambda_2^Y[\theta_Z(\beta_Y(T-t),\gamma_Y(T-t),0)-1]+\lambda_1^m[\theta_Z(0,0,\delta_Y(T-t))-1]=0
\end{aligned}
$$

$\forall t\in(0,T]$ with boundary conditions $\alpha_Y(0)=0$, $\beta_Y(0)=\omega$, $\gamma_Y(0)=0$ and $\delta_Y(0)=0$.

[16] This relies on the fact that the Poisson processes driving the jumps in v and in m are independent.

3 Coefficients for the Fourier Cosine Expansion

Here we give the expression for U_n, the Fourier cosine transform of VIX options' payoff. To ease notation, we define $\omega_n := \frac{n\pi}{b-a}$. U_n is given by:

$$U_n = \int_a^b \left(\sqrt{x} - K\right)^+ cos\left(\omega_n(x-a)\right) dx$$
$$= \frac{2}{b-a} Re\left\{ e^{-i\omega_n a}\left[\sqrt{b}\frac{e^{-i\omega_n b}}{i\omega_n} + \frac{\sqrt{\pi}}{2}\frac{1}{(-i\omega_n)^{3/2}}\left(\text{erfz}(\sqrt{-i\omega_n b}) - \text{erfz}(K\sqrt{-i\omega_n})\right)\right]\right\} \quad \text{for} \quad n > 0, \tag{39}$$

$$U_0 = \frac{2}{b-a}\left[\frac{2}{3}b^{3/2} - Kb + \frac{1}{3}K^3\right]. \tag{40}$$

4 Interpolation procedure for the S&P 500 options dataset

The interpolation method we have used to find a parametric fit for each slice of S&P 500 implied volatilities is developed in detail in Rebonato and Cardoso (2004). Here we give the main idea and the particular choice of parameters we have made so that the slices are well fitted.

The idea of Rebonato and Cardoso (2004) is to use a mixture of log-normal densities for the Futures price.[17] Log-normal densities mixtures can fit various different shapes including multimodal densities arising from jumps in the Futures price process. The ability of the method to recover any type of density (or equivalently smile) is well documented in Rebonato and Cardoso (2004) and chapter 9 of Rebonato (2004).

The attractive feature of this parametric representation for the density of the Futures price is that the pricing of call/put options can be performed using a mixture of Black-Scholes price. Additionally, the no-arbitrage condition is simply a condition on the expectation for the mixture of the Futures price.

In practice, a mixture of 4 log-normal densities is enough to have a nearly perfect fit. We minimize the Euclidean distance between market and mixture option prices using the CMA-ES algorithm mentioned in section 5.1. We have checked that the resulting fits are satisfactory by computing different measures of the distance between the market and model implied volatility slices. For instance, the RMSE between implied volatilities of the parametric fit and the true implied volatilities is most of the time below 0.25%. Sometimes the RMSE is larger and goes up to 2%, however this is not due to the inability of the mixture of log-normals to fit an implied volatility slice but due to the shape of the data. This is best explained by looking at a typical fit as displayed on Figure 6. We can see that the fit is nearly perfect, however the RMSE is not so close to zero because the input data is too rough. This phenomenon is amplified if the data has a sawtooth pattern (potential arbitrage) although the fit is very good.

Finally, using the parametric fit, we can sample "market option prices" for the desired strikes and moneynesses. We have chosen to re-sample option prices from each parametric slice uniformly in strike. We also choose not to resample options for which the strike is smaller than 40% or larger than 140% of the current Futures price. The reason is that there are usually only one or two options outside this interval of moneyness and we do not wish to re-sample options where the interpolation results could be driven by an outlier.

[17] They use a mixture for the stock price density but it is simple to adapt it to the Futures price when interest rates are constant.

5 Particle Filter

The filtering procedure consists in approximating the distribution $\hat{p}(L_t|y^t, \Theta, \mathcal{M})$ of unobservable (latent) factors L_t at every point in time t from available observations $y^t = \{y_1, ..., y_t\}$ and assuming that on one hand the model $\mathcal{M}$ is well specified and on the other hand its parameters Θ are known. In the remaining of this section we drop the dependence on the parameter set Θ and the model specification $\mathcal{M}$. The algorithm can be decomposed into the following steps.

Step 1: Initialization

Let us consider a time grid $\{0, 1, ..., t-1, t+1, ..., T\}$ in which we have available market measurements (e.g., S&P 500 forward returns, VIX levels and option prices). The initialization phase aims at simulating np possible initial values for the latent factors whose value depends on previous time $t-1$. We simulate np particles $\{v_0^{(i)}, m_0^{(i)}\}_{i=1,...,np}$ that are compatible with the initial value of the VIX squared. The next steps will be repeated for every time in the grid from 0 to $T-1$.

Step 2: First-stage resampling

At this stage we assume that we have np particles, which are possible values of m and v at time t. The goal is to assign weights (namely first-stage weights), which are proportional to the likelihood of market observations at time $t+1$ given the value of the particles at time t. Intuitively, particles that are compatible with the new observations will be assigned larger weights than other particles. For speed reasons, we do not consider options as part of the observation set in this step, but limit ourselves to the values of the indices.

The first-stage weight $\omega_{t+1}^{(i)}$ assigned to the i^{th} particle at time $t+1$ is given by:

$$\omega_{t+1}^{(i)} \propto p(y_{t+1}|L_t^{(i)})$$

where $p(y_{t+1}|L_t^{(i)})$ is the density of the observation vector y_{t+1} given the values of the particle vector $L_t^{(i)}$. The importance weights $\omega_{t+1}^{(i)}$ add up to 1 so that they define a proper probability distribution. Conditioning with respect to the number of jumps in Y, v and m gives:

$$\omega_{t+1}^{(i)} \propto \sum_{j,k,l} p(y_{t+1}|L_t^{(i)}, \Delta N_t^Y, \Delta N_t^v, \Delta N_t^m)\mathbb{P}(\Delta N_t^Y = j, \Delta N_t^v = k, \Delta N_t^m = l).$$

We make the assumption that observations are independent and therefore that the density can be expressed as the product of densities of forward S&P 500 log-returns and VIX levels. We do not augment the time space as we have daily observations as in Johannes, Polson, and Stroud (2009).

The next steps will consist in proposing new values for the state variables, generated from the mixture density:

$$\frac{\sum_{i=1}^{np} \omega_{t+1}^{(i)} q(.|L_t^{(i)})}{\sum_{i=1}^{np} \omega_{t+1}^{(i)}}.$$

We first resample particles according to a stratified resampling scheme:[18]

$$z(i) \sim StratRes(np, \omega_{t+1}^{(1)}, ..., \omega_{t+1}^{(np)}).$$

This allows to sample new values of the latent factors assuming equally weighted particles. Particles representing m_t and v_t are shuffled into a new set of particles: $\{m^{(j)}, v^{(j)}\}_{j=1..np} = \{m^{z(i)}, v^{z(i)}\}_{i=1..np}$. The new number of particles does not need to be equal to the old one. This resampling step eliminates all the particles that have no chance being valid given the new observations.

Step 3: Generating the jumps

The next step of the particle filter consists in propagating the latent factors according to their conditional density given the previous values and the new observations:

$$L_{t+1}^{(i)} \sim p(L_{t+1}|L_t^{(i)}, y_{t+1}).$$

Because the distribution $p(L_{t+1}|L_t^{(i)}, y_{t+1})$ is not known in closed-form, we use a proposal density $q(L_{t+1}^{(i)}|L_t^{(i)}, y_{t+1})$. Propagating v and m requires preliminary knowledge on the jump components.

We calculate the joint probability of jumps in Y, v and m between t and $t+1$ by:

$$\mathbb{P}(\Delta N_t^Y, \Delta N_t^v, \Delta N_t^m|y_{t+1}) \propto p(y_{t+1}|\Delta N_t^Y, \Delta N_t^v, \Delta N_t^m)\mathbb{P}(\Delta N_t^Y, \Delta N_t^v, \Delta N_t^m). \tag{41}$$

Conditionally on the jump sizes in v and m, the first part of the right hand-side is the density of a bivariate normal distribution. We assume that the exponential distribution can be expressed as a sum of Dirac functions at some chosen quantiles. Using Bayes' rule, we infer an approximation of $p(y_{t+1}|\Delta N_t^Y, \Delta N_t^v, \Delta N_t^m)$.

In order to better represent the tail of the distribution, we floor the probability of no jump in any process and simulate from the resulting distribution function the jumps in every process. We infer the total jump size following Johannes, Polson, and Stroud (2009). It has a normal distribution

[18] We checked that using a multinomial resampling scheme does not impact the overall results

$p(\Delta J_t^Y | \Delta N_t^Y, y_{t+1})$: $\Delta J_t^{Y(i)} \sim \mathcal{N}(\mu_J^{Y(i)}, \sigma_J^{Y(i)})$ where $\mu_J^{Y(i)}$ and $\sigma_J^{Y(i)}$ are given by:

$$(\sigma_J^{Y(i)})^2 = \left(\frac{1}{\Delta N_t^{Y(i)} \sigma_Y^2} + \frac{1}{\hat{v}_{t+1}} \right)^{-1}$$
$$\mu_J^{Y(i)} = (\sigma_J^{Y(i)})^2 \frac{Y_{t+1} - \tilde{\mu}_Y}{\hat{v}_{t+1}} + \frac{(\sigma_J^{Y(i)})^2}{\sigma_Y^2} \mu_Y$$

where $\hat{v}_{t+1}$ is an estimate of v_{t+1} given the information we have up to time t; we use $\hat{v}_{t+1}^{(i)} = \mathbb{E}[v_{t+1}|v_t]$ and

$$\tilde{\mu}_Y = -(\lambda^{Y(\mathbb{Q})}(\theta_Z(1,0,0)^{\mathbb{Q}} - 1) + \frac{1}{2} v_t - \Delta N_t^{Y(i)})\Delta_t.$$

Propagating the latent factors

The latent factors v and m are propagated following a Milstein discretization scheme of the SDE. We use the Full Truncation method to prevent them from taking negative values.

Filtering latent factors

Weights called 'second-stage weights' are assigned to the particle sets and provide a non-biased estimation of the distribution of latent factors. The most-likely value every factor is taken to be the expectation of the estimated density function for each factor. These weights denoted by $\pi_{t+1}^{(i)}$ are proportional to the likelihood of observations at time $t+1$ given the propagated particles $L_{t+1}^{(i)}$, and corrected for the pre-weighting:

$$\pi_{t+1}^{(i)} \propto \frac{p(L_{t+1}^{(i)}|L_t^{(i)}) p(y_{t+1}|L_{t+1}^{(i)})}{\omega_{t+1}^{z(i)} q(L_{t+1}^{(i)}|L_t^{(i)}, y_{t+1})}.$$

The posterior distribution of state variables is approximated by:

$$\hat{p}(L_{t+1}|y^{t+1}) = \sum_{i=1}^{np} \pi_{t+1}^{(i)} \delta(L_{t+1} - L_{t+1}^{(i)}).$$

Maximum Likelihood

The algorithm described above holds if one assumes that the model parameters are known. However, Pitt (2002) builds on Gordon, Salmond and Smith (1993) to show that the parameters can be estimated using the Maximum Likelihood Importance Sampling Criterion, defined as the product over time of the averages of the second-stage weights. The likelihood of observations given the values of particles is then estimated by the average of the second-stage weights over particles:

$$\hat{p}(y^n|\Theta) = \prod_{t=2}^{n} \hat{p}(y_t|y^{t-1}, \Theta)\hat{y_1|}\Theta$$

where

$$\hat{p}(y_t|y^{t-1}, \Theta) = \frac{1}{np} \sum_{i=1}^{np} \pi_t^{(i)}.$$

6 Tables and Figures

6.1 Tables

Table 1: Descriptive Statistics for S&P 500 Futures log-returns and VIX levels

	March 2006 - November 2008				December 2008 - October 2010			
	Mean	Standard dev.	Skewness	Kurtosis	Mean	Standard dev.	Skewness	Kurtosis
Log-returns SPX	-0.0007	0.0159	0.0722	14.0772	0.0007	0.0158	-0.3283	6.8683
VIX	0.2044	0.1208	2.6560	10.9620	0.2907	0.1025	1.1929	3.8631

Table 2: Quantiles of implied volatility values for S&P 500 and VIX options. The quantiles are calculated using all maturities and moneyness available from March 2006 to October 2010.

Quantiles	SPX IVs	VIX IVs
25%	16.93%	62.95%
50%	23.16%	75.03%
75%	32.77%	91.68%

Table 3: Root Mean Squared Error in implied volatilities when calibrating the SVJ and the SVJ2 models to S&P 500 and VIX options on four different dates. On each date, the distance (23) is minimized using a global optimizer over the model parameters under $\mathbb{Q}$. On each date, the data is composed of all S&P 500 options implied volatilities and all VIX implied volatilities together. Here we report the distances $RMSE\_SPX$ and $RMSE\_VIX$ but the minimization is run over $RMSE = \frac{1}{2}(RMSE\_SPX + RMSE\_VIX)$.

RMSE (%)	20070711		20081022		20090610		20100505	
	SPX	VIX	SPX	VIX	SPX	VIX	SPX	VIX
SVJ	1.766	8.007	2.508	13.601	1.870	9.997	1.271	11.599
SVJ2	0.852	3.885	2.427	12.757	2.110	7.933	1.169	5.153

Table 4: Root Mean Square Pricing Errors (on SPX and VIX options) for the SVJ2, the SVJ and the SV2 models depending on the dataset these models were estimated on, and conditional on the moneyness and maturity of options. The last column with heading # describes the number of options in each category given in the first column.

	Dataset 1			Dataset 2			Dataset 3			Dataset 4			
	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	#
Fitting SPX options in the in-sample period													
Overall	11.873	6.999	9.128	2.510	6.874	3.985	6.875	4.754	6.281	3.068	5.697	3.634	14351
Deep OTM puts	2.902	4.423	3.680	2.459	4.117	3.335	1.614	2.086	2.715	2.926	3.784	2.761	1365
OTM puts	6.286	7.203	5.283	2.560	6.892	4.219	2.559	3.645	2.458	3.084	5.940	3.499	6498
Options close ATM	13.604	7.731	10.880	2.822	7.501	3.964	5.672	4.835	5.874	3.374	6.177	4.444	3673
OTM calls	19.631	6.845	14.456	1.895	7.349	3.666	11.629	5.900	11.217	2.622	5.338	3.179	2287
Deep OTM calls	19.154	4.894	12.559	2.008	5.511	4.033	19.692	11.225	13.923	2.758	4.718	2.734	528
Short mat options	2.619	2.073	2.457	1.183	1.587	1.265	2.600	2.945	2.753	1.412	1.843	1.421	4476
Medium mat options	7.856	5.384	6.428	2.086	5.017	2.455	5.339	4.092	5.263	2.452	4.324	3.345	5499
Long mat options	19.436	10.970	14.669	3.710	10.988	6.547	10.595	6.648	9.317	4.611	8.915	5.214	4376
Fitting SPX options in the out-of-sample period													
Overall	9.974	11.030	9.330	3.292	11.074	6.137	10.973	6.405	9.203	3.881	8.950	3.881	13323
Deep OTM puts	3.639	5.892	4.778	3.005	5.591	4.300	1.845	3.998	2.981	3.531	5.375	3.218	3046
OTM puts	6.167	13.221	8.121	3.875	12.483	7.334	4.805	7.393	4.373	4.907	10.783	4.893	4755
Options close ATM	11.373	13.220	10.730	3.301	13.484	6.452	11.094	7.539	10.300	3.605	10.171	3.927	1962
OTM calls	15.317	10.725	12.873	2.629	12.126	5.671	16.807	6.465	14.604	2.616	8.578	2.758	2349
Deep OTM calls	15.723	7.729	11.548	2.540	8.890	5.174	21.866	4.831	15.770	2.351	6.460	2.426	1211
Short mat options	2.732	2.219	2.609	1.545	1.940	1.564	3.009	3.765	3.442	1.564	2.013	1.510	3580
Medium mat options	7.158	7.195	6.753	2.292	6.829	3.368	7.871	3.964	7.106	2.537	5.446	2.717	5678
Long mat options	15.745	18.015	14.684	5.106	18.262	10.268	17.323	10.001	14.021	6.183	14.749	6.086	4065
Fitting VIX options in the in-sample period													
Overall	0.021	0.026	0.022	0.017	0.025	0.019	0.010	0.015	0.013	0.016	0.021	0.016	4148
Deep OTM options	0.008	0.007	0.007	0.006	0.007	0.006	0.004	0.004	0.005	0.006	0.007	0.006	1100
OTM options	0.014	0.013	0.012	0.011	0.012	0.011	0.004	0.007	0.007	0.010	0.013	0.010	963
Options close ATM	0.021	0.020	0.018	0.015	0.020	0.015	0.005	0.013	0.009	0.013	0.018	0.013	1075
ITM options	0.032	0.046	0.039	0.029	0.044	0.033	0.019	0.027	0.023	0.028	0.035	0.028	1010
Short mat options	0.017	0.022	0.020	0.013	0.021	0.016	0.011	0.015	0.012	0.014	0.019	0.014	1948
Medium mat options	0.023	0.029	0.024	0.020	0.028	0.022	0.009	0.015	0.013	0.018	0.023	0.018	2200
Fitting VIX options in the out-of-sample period													
Overall	0.025	0.039	0.032	0.021	0.039	0.027	0.014	0.024	0.017	0.023	0.037	0.022	5118
Deep OTM options	0.007	0.009	0.009	0.009	0.009	0.009	0.006	0.006	0.007	0.009	0.010	0.008	2039
OTM options	0.014	0.020	0.018	0.016	0.020	0.017	0.006	0.014	0.011	0.016	0.021	0.014	916
Options close ATM	0.025	0.035	0.029	0.022	0.035	0.025	0.009	0.025	0.016	0.023	0.035	0.021	1109
ITM options	0.046	0.076	0.059	0.036	0.074	0.051	0.027	0.044	0.032	0.041	0.069	0.040	1054
Short mat options	0.022	0.032	0.028	0.017	0.032	0.022	0.010	0.017	0.014	0.020	0.030	0.019	1054
Medium mat options	0.027	0.044	0.034	0.024	0.043	0.031	0.016	0.029	0.019	0.026	0.041	0.024	2237

Table 5: Estimated parameters from the particle filter for the different models and datasets.

	Dataset 1			Dataset 2			Dataset 3			Dataset 4		
	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2
					$\mathbb{P}$ and $\mathbb{Q}$ parameters							
λ_0^{Yv}	0.394	0.030	0	2.722	0.000	0	0.074	0.030	0	0.065	0.080	0
λ_1^{Yv}	1.650	0.007	0	0.079	0.212	0	0.122	1.649	0	0.105	0.990	0
λ_2^{Yv}	2.438	0	0	0.106	0	0	1.745	0	0	0.066	0	0
λ_0^{m}	0.134	0	0	0.080	0	0	0.267	0	0	0.267	0	0
λ_1^{m}	0.070	0	0	0.644	0	0	0.023	0	0	0.554	0	0
σ_m	0.150	0	0.147	0.113	0	0.216	0.163	0	0.209	0.185	0	0.266
σ_v	0.715	0.718	0.717	0.951	0.452	0.871	0.730	0.352	0.765	0.763	0.427	0.861
ρ_{Yv}	-0.550	-0.625	-0.578	-0.910	-0.925	-0.792	-0.645	-0.777	-0.446	-0.859	-0.807	-0.754
m_0		0.027			0.029			0.023			0.028	
					$\mathbb{P}$ parameters							
κ_v^P	10.002	11.057	9.723	7.608	25.690	0.262	7.840	7.389	8.231	6.821	8.935	8.549
κ_m^P	0.060	0	0.057	0.212	0	0.399	1.247	0	0.452	1.954	0	0.924
θ_m^P	0.501	0	0.504	0.036	0	0.077	0.014	0	0.049	0.011	0	0.039
ν_m^P	0.002	0	0	0.127	0	0	0.111	0	0	0.009	0	0
ν_v^P	0.112	0.155	0	0.265	0.079	0	0.121	0.030	0	0.169	0.247	0
μ_Y^P	-0.060	0.150	0	-0.209	0.105	0	-0.088	0.150	0	-0.060	-0.010	0
σ_Y^P	0.041	0.007	0	0.475	0.998	0	0.125	0.011	0	0.039	0.020	0
η_Y	0.721	0.000	0.733	0.150	-10.165	0.643	0.613	0.330	0.450	0.740	0.860	0.850
					$\mathbb{Q}$ parameters							
κ_v^Q	10.002	9.799	10.379	8.022	3.537	7.057	7.835	3.669	7.403	6.554	4.807	5.380
κ_m^Q	0.111	0	0.108	0.383	0	1.674	0.158	0	0.169	0.819	0	0.361
θ_m^Q	0.272	0	0.266	0.020	0	0.018	0.110	0	0.130	0.026	0	0.101
ν_m^Q	0.114	0	0	0.020	0	0	0.006	0	0	0.011	0	0
ν_v^Q	0.157	0.763	0	0.010	1.583	0	0.326	0.472	0	0.109	0.057	0
μ_Y^Q	0.040	-0.120	0	0.020	0.162	0	0.133	-0.090	0	-0.090	-0.140	0
σ_Y^Q	0.033	0.006	0	0.006	0.000	0	0.064	0.013	0	0.148	0.100	0
Log-likelihood	10626	10572	10520	9785	4170	8205	9893	8220	9123	4882	-2749	2239
AIC	-21206	-21116	-21020	-19518	-8306	-16384	-19730	-16402	-18216	-9702	5542	-4442
BIC	-21103	-21053	-20975	-19402	-8230	-16326	-19605	-16317	-18149	-9563	5641	-4361

Table 6: Statistics on in-sample and out-sample errors on VIX levels, for the different models and datasets.

	Dataset 1			Dataset 2			Dataset 3			Dataset 4		
	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2
In-sample fitting levels												
Mean error VIX	-0.001	-0.001	-0.001	0.001	0.002	0.001	0.000	-0.001	0.001	0.001	0.003	0.002
RMSE VIX	0.023	0.023	0.022	0.022	0.020	0.021	0.022	0.020	0.023	0.021	0.020	0.022
Out-of-sample fitting levels												
Mean error VIX	0.001	0.001	0.001	0.001	0.002	0.002	-0.001	-0.005	-0.001	0.002	0.003	0.003
RMSE VIX	0.022	0.022	0.021	0.023	0.023	0.024	0.020	0.022	0.021	0.023	0.023	0.024

Table 7: Diebold-Mariano test values for in-sample and out-sample errors on VIX levels, for different models calibrated to log-returns and VIX levels. Two loss functions are considered namely the absolute and quadratic error. Standard errors are calculated using the Newey and West (1987) estimator of the standard deviation of the error, which takes into account possible autocorrelation and heteroscedasticity of the time-series. The number of lags is optimally chosen following Andrews (1991).

	SVJ		SV2	
	Abs loss function	Quad loss function	Abs loss function	Quad loss function
In-sample	-0.278	-0.001	-0.253	-0.001
Out-of-sample	-0.243	-0.002	-0.253	-0.002

Table 8: Root Mean Square Relative Pricing Errors (on SPX and VIX options) for the SVJ2, the SVJ and the SV2 models depending on the dataset these models were estimated on, and conditional on the moneyness and maturity of options. The last column with heading # describes the number of options in each category given in the first column.

	Dataset 1			Dataset 2			Dataset 3			Dataset 4			
	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	SVJ2	SVJ	SV2	#
Fitting SPX options in the in-sample period													
Overall	2.946	0.850	1.849	0.249	0.387	0.271	1.993	0.954	1.561	0.261	0.420	0.304	14351
Deep OTM puts	0.738	0.625	0.573	0.366	0.592	0.421	0.444	0.655	0.522	0.392	0.618	0.465	1365
OTM puts	0.501	0.398	0.449	0.241	0.352	0.275	0.404	0.320	0.414	0.210	0.354	0.277	6498
Options close ATM	0.675	0.367	0.484	0.166	0.223	0.168	0.433	0.427	0.343	0.227	0.227	0.213	3673
OTM calls	5.221	1.779	3.246	0.262	0.440	0.251	3.564	1.962	2.676	0.278	0.522	0.320	2287
Deep OTM calls	10.500	1.396	6.505	0.370	0.680	0.369	7.012	2.100	5.617	0.471	0.861	0.508	528
Short mat options	1.785	0.979	1.151	0.327	0.391	0.347	1.689	1.345	1.425	0.312	0.432	0.359	4476
Medium mat options	2.798	0.797	1.725	0.206	0.341	0.217	2.016	0.770	1.527	0.230	0.365	0.242	5499
Long mat options	3.922	0.767	2.473	0.202	0.435	0.242	2.237	0.622	1.727	0.240	0.471	0.314	4376
Fitting SPX options in the out-of-sample period													
Overall	2.337	0.751	1.631	0.402	0.623	0.447	2.638	1.092	2.219	0.420	0.609	0.440	13323
Deep OTM puts	0.729	0.898	0.812	0.612	0.860	0.718	0.664	0.666	0.716	0.678	0.880	0.648	3046
OTM puts	0.455	0.590	0.482	0.338	0.530	0.407	0.439	0.412	0.431	0.350	0.512	0.386	4755
Options close ATM	0.307	0.379	0.298	0.173	0.326	0.160	0.256	0.532	0.297	0.156	0.297	0.166	1962
OTM calls	2.744	1.006	1.988	0.281	0.570	0.260	3.028	1.908	2.560	0.248	0.496	0.255	2349
Deep OTM calls	6.571	0.772	4.344	0.429	0.711	0.282	7.537	1.952	6.270	0.375	0.693	0.564	1211
Short mat options	1.427	0.992	1.096	0.486	0.600	0.514	1.852	1.668	1.595	0.482	0.648	0.517	3580
Medium mat options	2.447	0.609	1.633	0.354	0.587	0.406	2.933	0.918	2.405	0.394	0.564	0.391	5678
Long mat options	2.783	0.674	1.982	0.381	0.688	0.438	2.787	0.526	2.413	0.395	0.633	0.431	4065
Fitting VIX options in the in-sample period													
Overall	1.088	0.738	0.795	0.530	0.627	0.587	0.237	0.380	0.310	0.453	0.677	0.514	4148
Deep OTM options	1.309	0.764	0.835	0.643	0.679	0.702	0.369	0.434	0.495	0.604	0.844	0.633	1100
OTM options	1.218	0.811	0.898	0.564	0.654	0.669	0.186	0.385	0.226	0.451	0.708	0.567	963
Options close ATM	1.054	0.751	0.867	0.457	0.596	0.517	0.125	0.351	0.188	0.356	0.590	0.434	1075
ITM options	0.629	0.609	0.514	0.420	0.571	0.407	0.179	0.341	0.203	0.340	0.506	0.380	1010
Short mat options	1.101	0.761	0.909	0.483	0.586	0.642	0.272	0.348	0.342	0.431	0.669	0.543	1948
Medium mat options	1.075	0.717	0.677	0.568	0.660	0.532	0.200	0.406	0.279	0.471	0.683	0.488	2200
Fitting VIX options in the out-of-sample period													
Overall	0.650	0.894	0.793	0.689	0.883	0.734	0.432	0.556	0.605	0.719	0.923	0.655	5118
Deep OTM options	0.714	0.916	0.946	0.881	0.903	0.901	0.598	0.483	0.853	0.913	0.993	0.842	2039
OTM options	0.652	0.915	0.756	0.649	0.902	0.711	0.297	0.635	0.435	0.695	0.955	0.601	916
Options close ATM	0.646	0.873	0.675	0.509	0.871	0.580	0.217	0.614	0.327	0.531	0.880	0.467	1109
ITM options	0.505	0.855	0.588	0.401	0.836	0.499	0.296	0.553	0.307	0.427	0.791	0.407	1054
Short mat options	0.666	0.846	0.798	0.658	0.835	0.702	0.414	0.487	0.619	0.695	0.891	0.651	1054
Medium mat options	0.637	0.930	0.789	0.712	0.918	0.757	0.445	0.605	0.595	0.738	0.948	0.657	2237

Table 9: Diebold-Mariano test values for in-sample and out-sample errors on VIX option prices, for the different models and estimation datasets. Two loss functions are considered namely the average Mean Square Error (MSE) and the average Mean Square Relative Error (MSRE). Standard errors are calculated using the Newey and West (1987) estimator of the standard deviation of the error, which takes into account possible autocorrelation and heteroscedasticity of the time-series. The number of lags is optimally chosen following Andrews (1991).

	Dataset 3				Dataset 4			
	SVJ		SV2		SVJ		SV2	
	MSE	MSRE	MSE	MSRE	MSE	MSRE	MSE	MSRE
In-sample	0.000	0.192	0.000	0.056	0.000	0.450	0.000	0.302
Out-of-sample	0.000	0.292	0.000	0.210	0.000	0.430	0.000	-0.552

Table 10: Diebold-Mariano test values for in-sample and out-sample errors on S&P 500 option prices, for different models and estimation datasets. Two loss functions are considered namely the average Mean Square Error (MSE) and the average Mean Square Relative Error (MSRE) . Standard errors are calculated using the Newey and West (1987) estimator of the standard deviation of the error, which takes into account possible autocorrelation and heteroscedasticity of the time-series. The number of lags is optimally chosen following Andrews (1991).

	Dataset 2				Dataset 4			
	SVJ		SV2		SVJ		SV2	
	MSE	MSRE	MSE	MSRE	MSE	MSRE	MSE	MSRE
In-sample	13.023	0.127	5.429	0.062	48.639	0.138	24.501	0.166
Out-of-sample	12.966	0.191	6.359	0.147	27.154	0.299	-0.161	0.126

6.2 Figures

Figure 1: Joint evolution of the VIX (dashed curve) and SPX index (solid curve) values from February 2006 to June 2010. The left y-axis corresponds to the VIX values (in percentage) and the right y-axis to the S&P 500 index values.

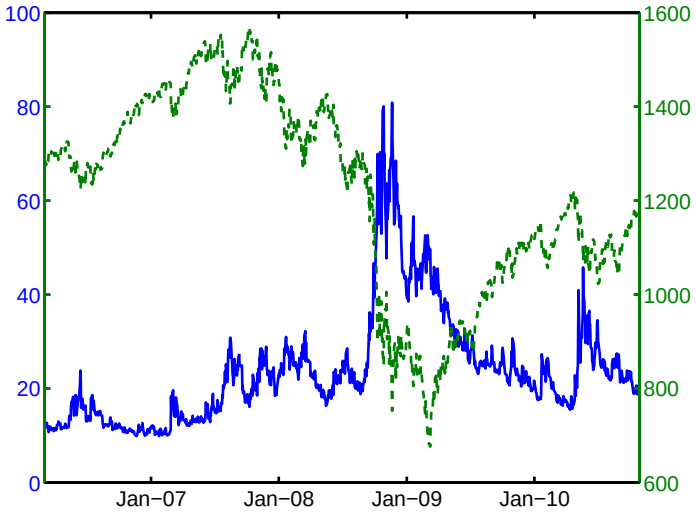


Figure 2: SPX and VIX implied volatility skews on four different dates as a function of log-moneyness ($\log K/F$). For each market, the scale is the same across days. The maturities T are quoted in years.

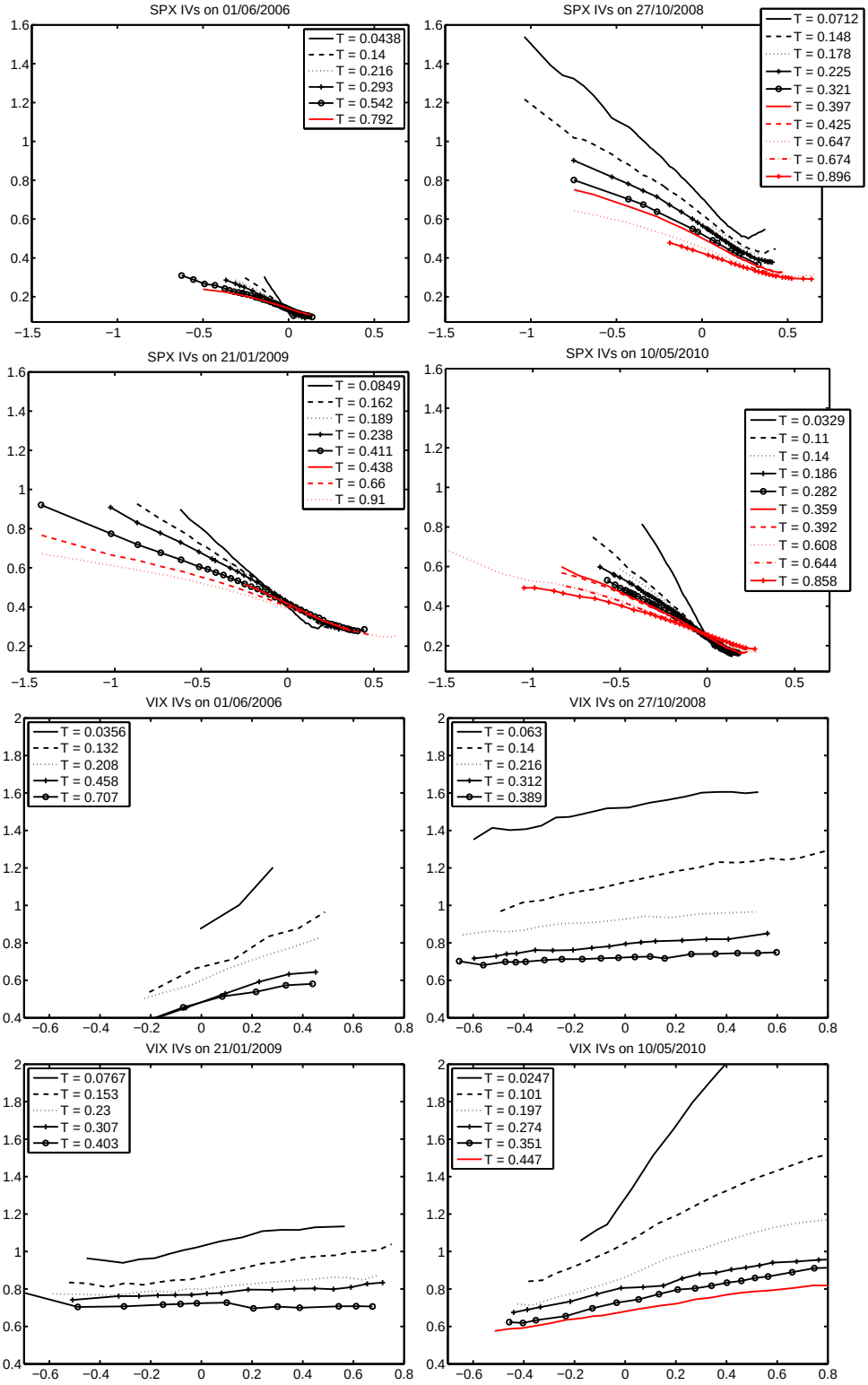


Figure 3: One month expected returns of the S&P 500 Futures (Panel A) and VIX index (Panel B) implied by S&P 500 and VIX options with maturity one month from March 1st, 2006 to October 29th, 2010. We use the method described in Bakshi, Kapadia, and Madan (2003). Returns are expressed in percentage units.

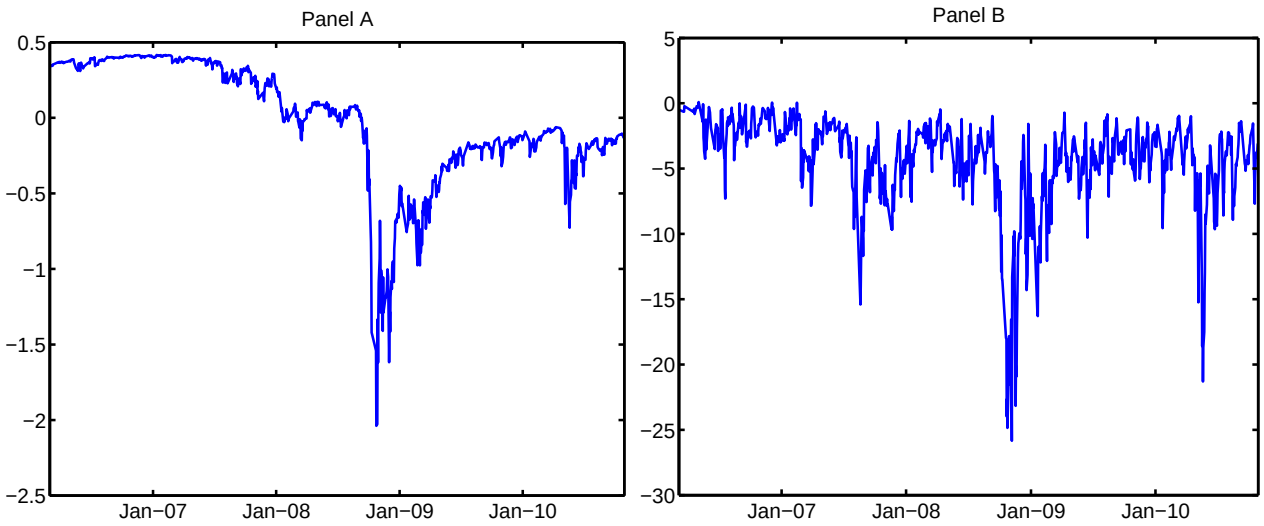


Figure 4: Market and Heston model implied volatilities for VIX options (four maturities) on October 20th, 2010 plotted with respect to forward log-moneyness ($\log K/F(T)$). The market (resp. model) implied volatilities are represented by the crosses (resp. the solid line). These fits are obtained by minimizing relative errors between market implied volatilities and the Heston model implied volatility.

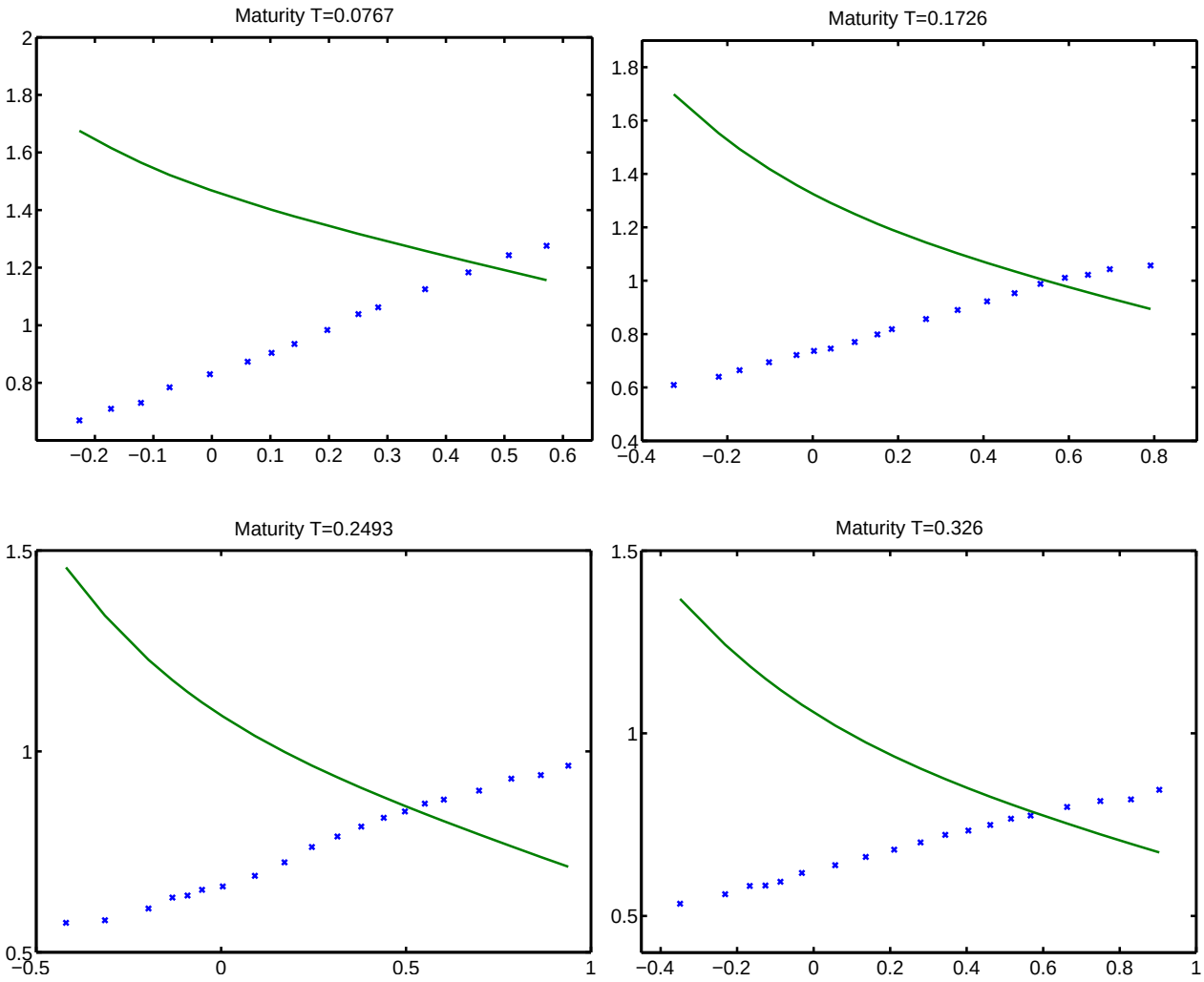


Figure 5: Comparative fit of 4 different maturities of VIX options for the SVJ2 (panles A to D) and the SVJ models (panels E to H) on July 11th, 2007. The crosses are the market implied volatilities and the curve represents the model volatilities. The implied volatilities are plotted as a function of forward log-moneyness ($\log K/F(T)$).

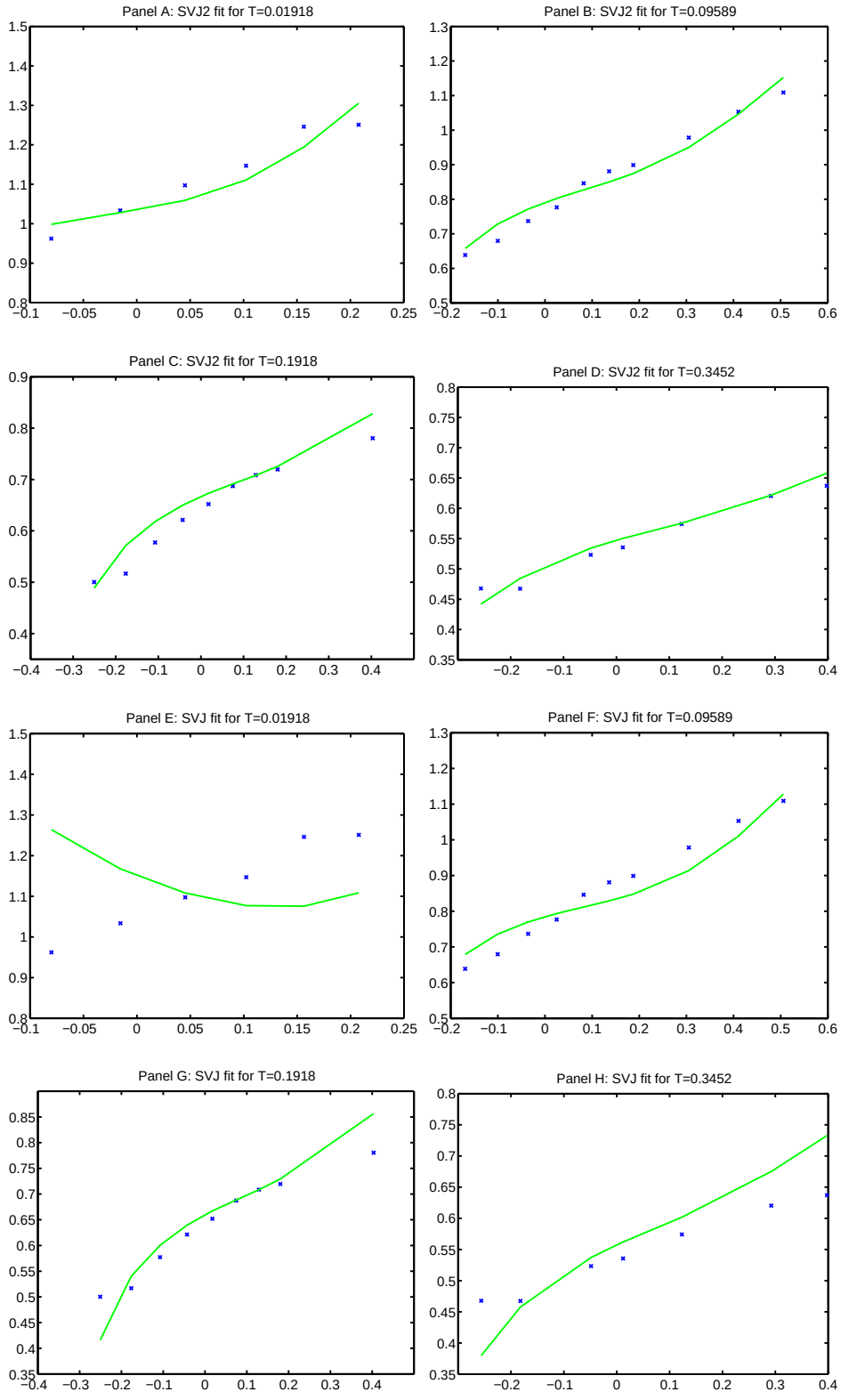


Figure 6: Typical interpolation of market implied volatilities (circles) using a mixture of log-normal densities for the density of Futures prices. The implied volatilities are plotted as a function of forward log-moneyness ($\log K/F(T)$).

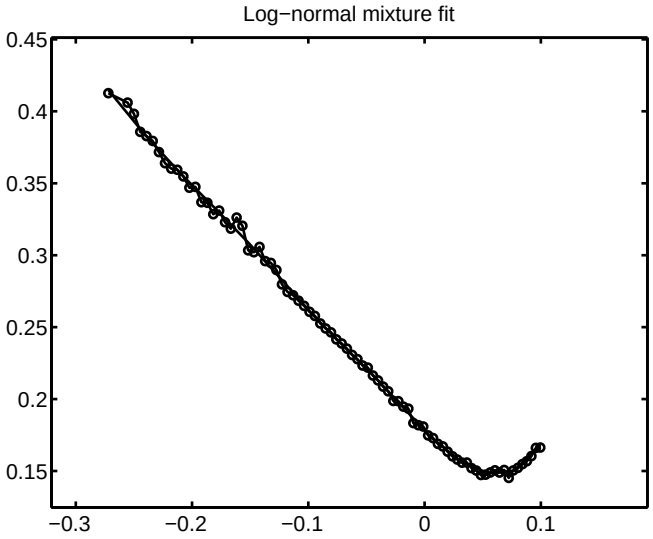


Figure 7: Filtered trajectories of the latent factor $\sqrt{v}$ when estimating the SVJ2 (solid line), the SVJ (dashed line) and the SV2 (dashed dotted line) models over the S&P 500 log-returns, the VIX index values, VIX and S&P 500 option prices from March 2006 to November 2008 (685 days). The shaded part of the graph represents the out-of-sample period, from December 2008 until October 2010.

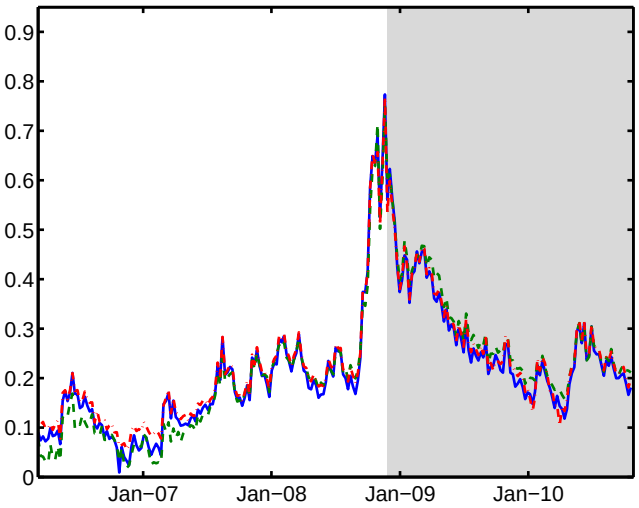


Figure 8: Difference between the filtered trajectory of the latent factor $\sqrt{v}$ when estimating the SVJ2 model over *Dataset 4* and other datasets. Panel A represents the difference with the filtered trajectory using *Dataset 1*. Panel B discloses the difference with the filtered trajectory using *Dataset 2* and Panel C displays the difference with the filtered trajectory using *Dataset 3*.

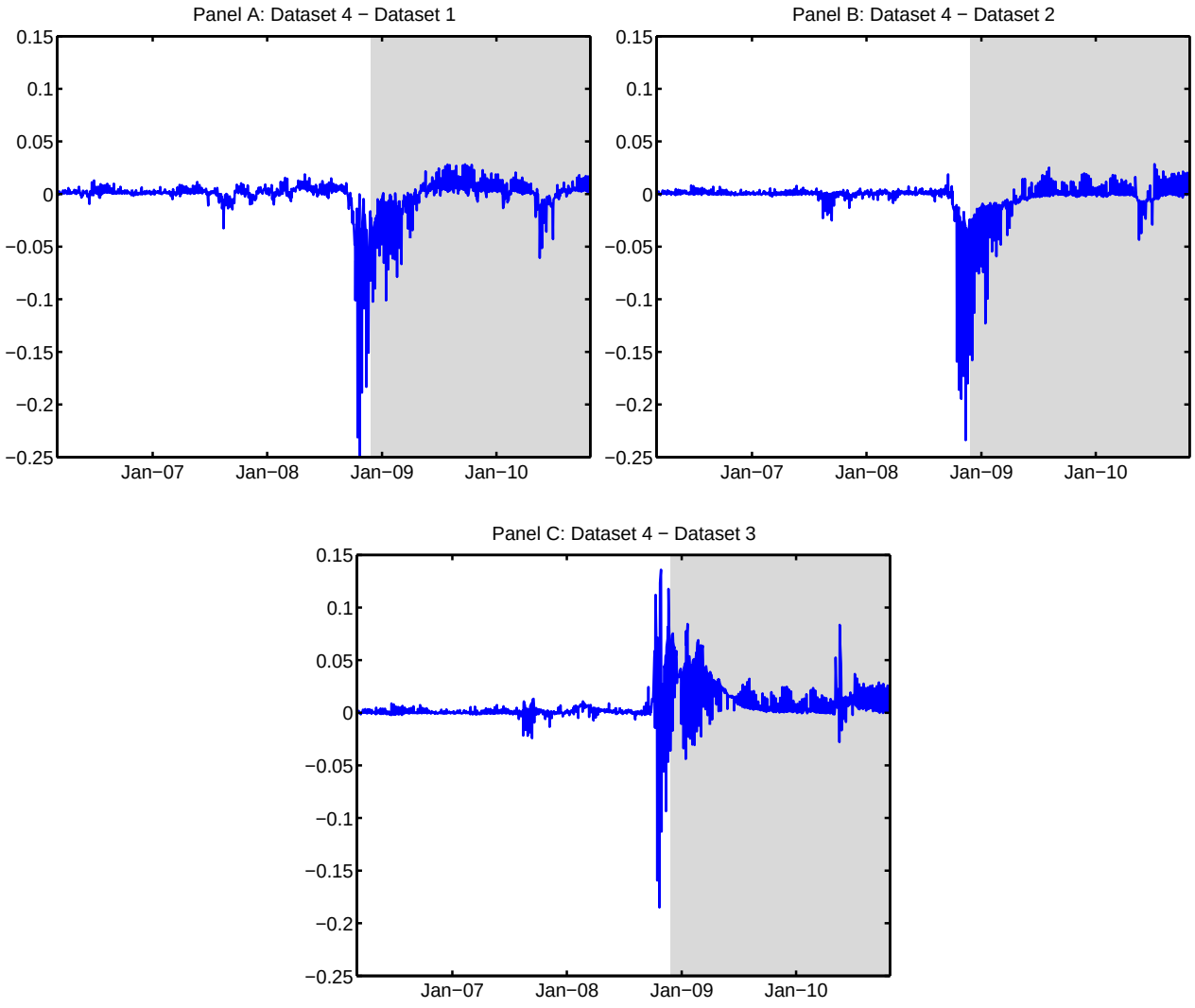


Figure 9: Filtered jump sizes in the variance process v when estimating the SVJ2 (solid line) and the SVJ (dashed line) models over the different datasets from March 2006 to November 2008 (685 days). We consider that there is a jump when the filtered probability of jump is larger than 50%. Panel A corresponds to *Dataset 1* which comprises the underlying forward returns on the S&P 500 and the VIX levels. Panel B corresponds to *Dataset 2* which consists of the underlying index levels plus S&P 500 options. Panel C corresponds to *Dataset 3* which comprises the underlying index levels plus VIX options. Finally Panel D corresponds to *Dataset 4* which gathers all data sources considered. The shaded part of the graph represents the out-of-sample period, from December 2008 until October 2010.

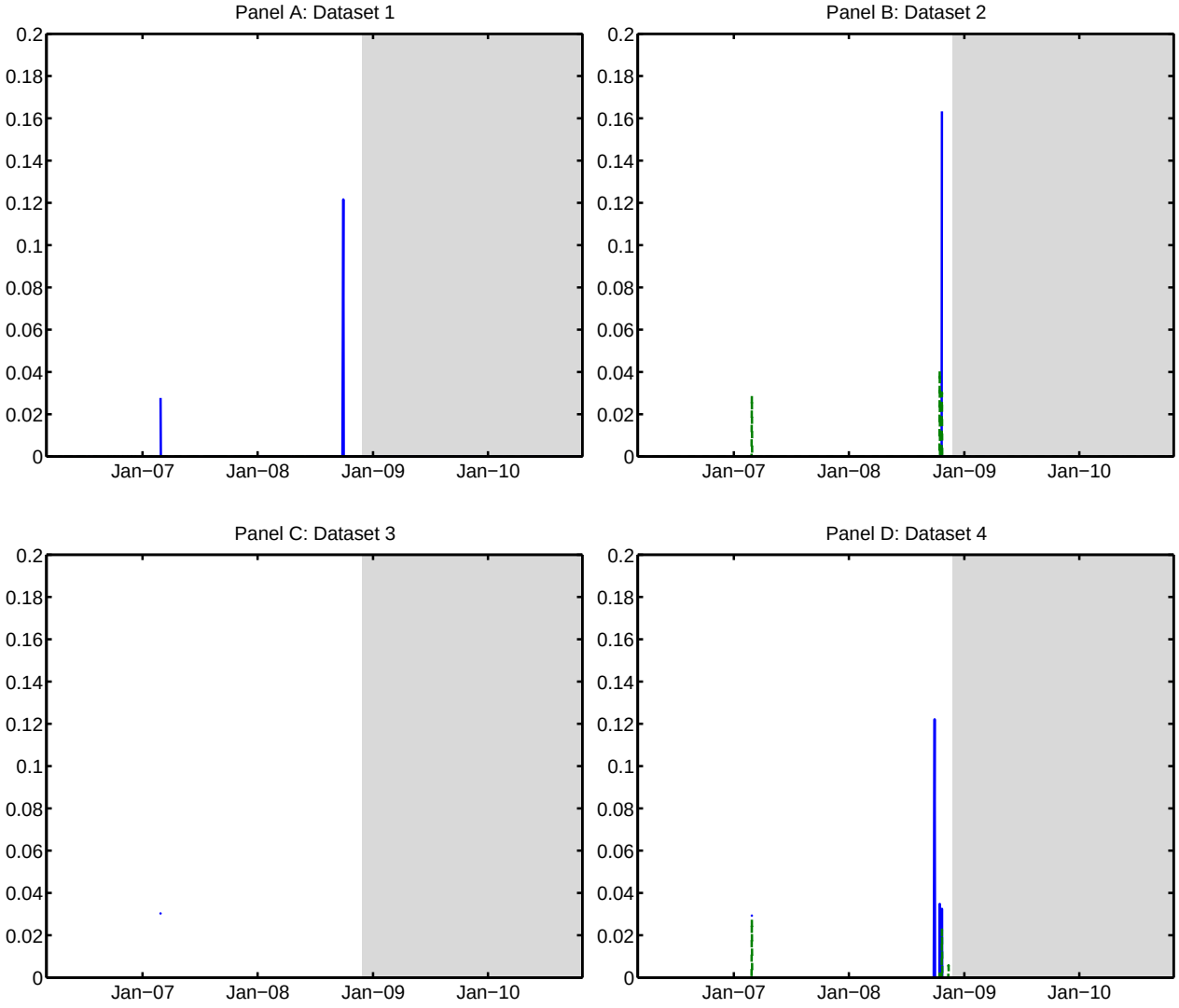


Figure 10: Filtered jump sizes in the process m when estimating the SVJ2 model over the datasets 3 and 4 from March 2006 to November 2008 (685 days). We consider that there is a jump when the filtered probability of jump is larger than 50%. The graphs correspond to *Dataset 2* (Panel A) and *Dataset 4* (Panel B). The shaded part of the graph represents the out-of-sample period, from December 2008 until October 2010.

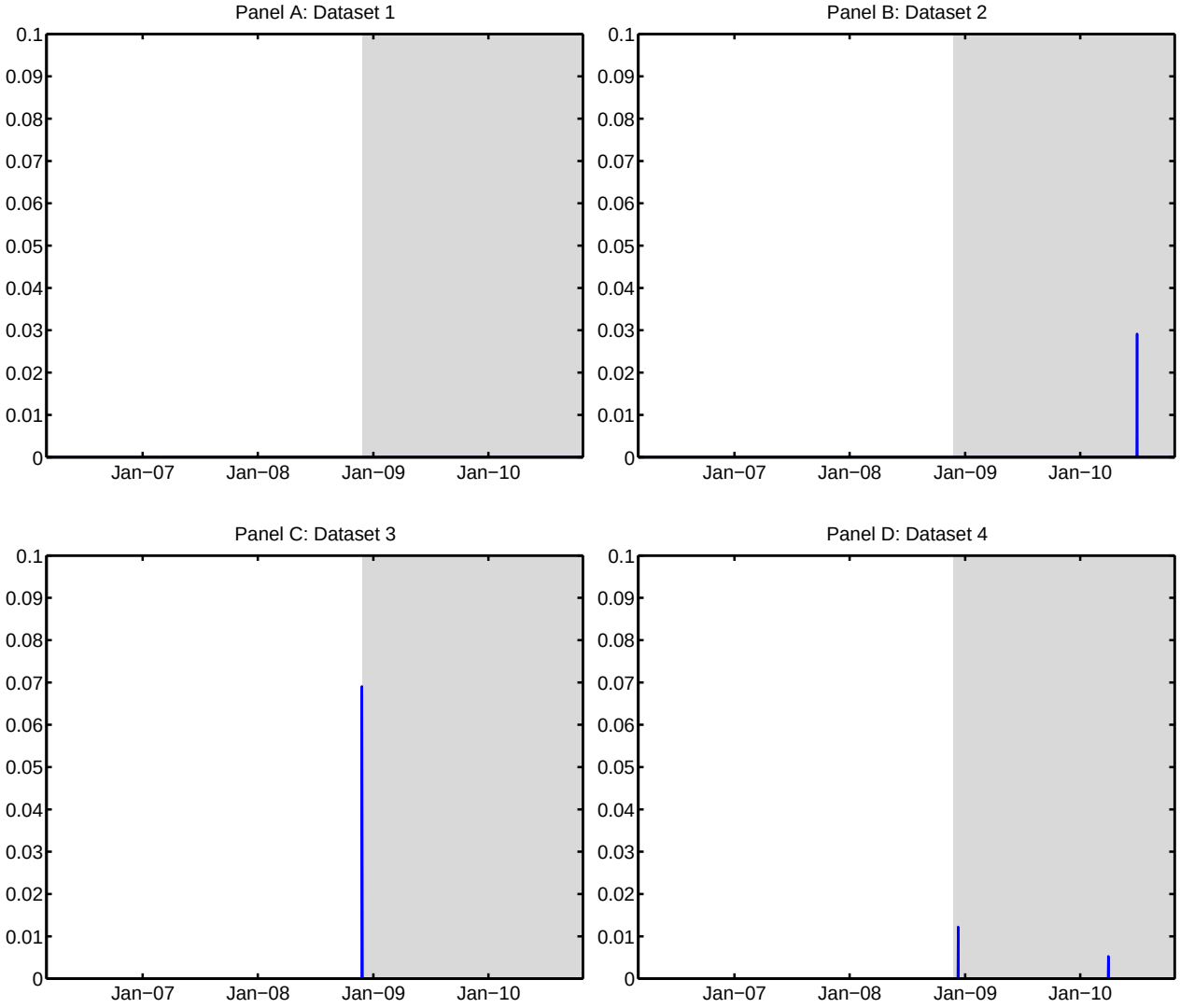


Figure 11: Filtered trajectories of the latent factor m when estimating the SVJ2 (solid line), SVJ (horizontal dashed line) and the SV2 (dashed dotted line) model over the different datasets from March 2006 to November 2008 (685 days). Panel C corresponds to *Dataset 3* which comprises the underlying index levels plus VIX options. Panel D corresponds to *Dataset 4* which gathers all data sources considered. The shaded part of the graph represents the out-of-sample period, from December 2008 until October 2010.

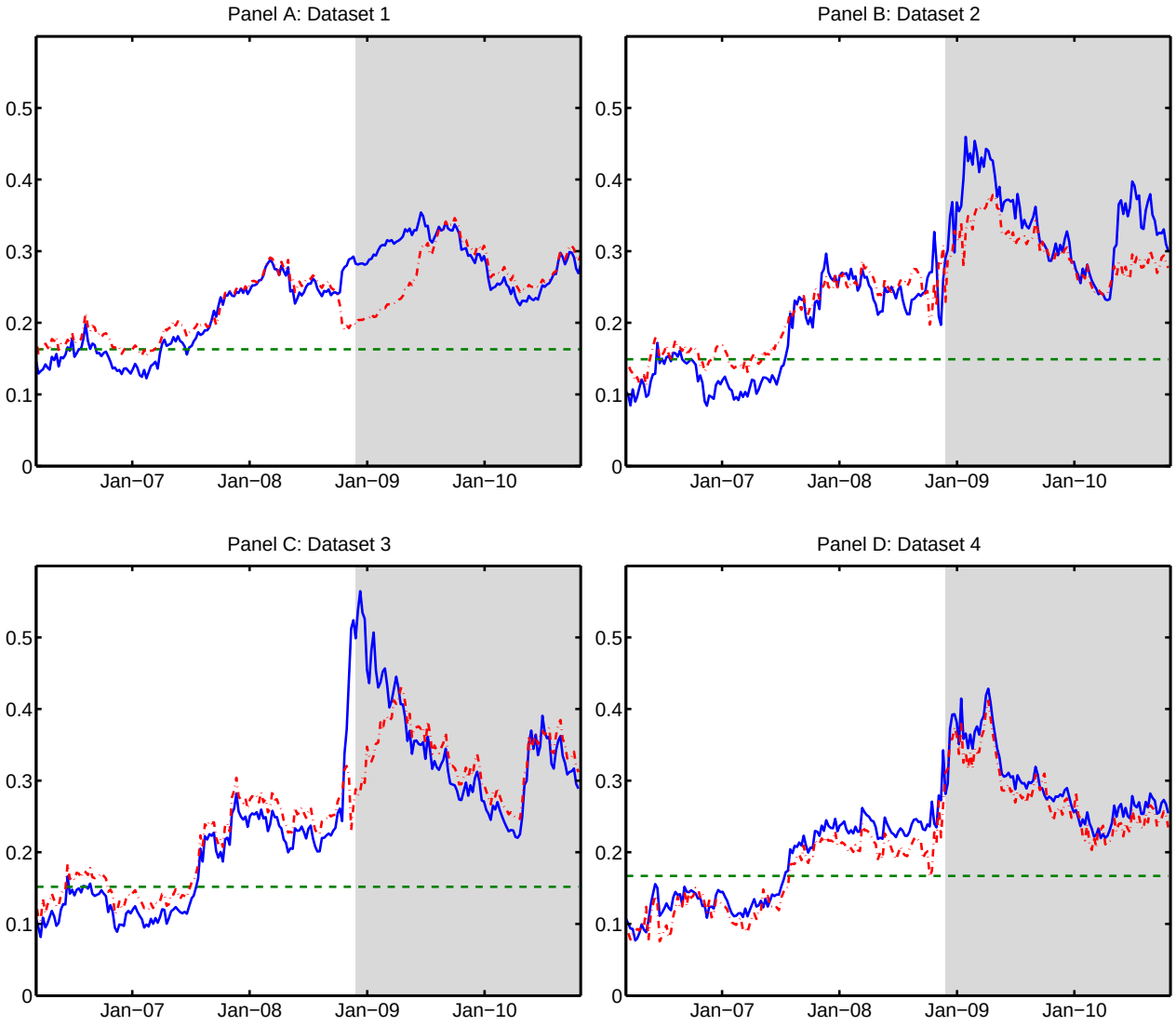


Figure 12: Fitting of VIX index values for the SVJ2 model when the model is calibrated to log-returns and VIX levels (*Dataset 1*) from March 2006 to November 2008 (685 days). The crosses represent market data, the line the filtered values. The shaded part of the graph represents the out-of-sample period, from December 2008 until October 2010.

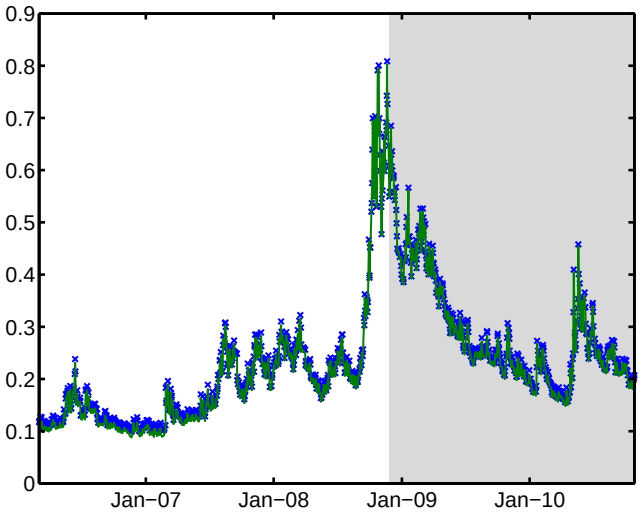


Figure 13: 1 month risk-neutral skewness and kurtosis of the distribution of returns implied by 1 month SPX options prices when estimating the SVJ2 (solid line), the SVJ (dashed line) and the SV2 (dashed dotted line) models over *Dataset 4* (indices as well as SPX and VIX options) from March 2006 to November 2008 (685 days). We use the method described in Bakshi, Kapadia, and Madan (2003). The shaded part of the graph represents the out-of-sample period, from December 2008 until October 2010.

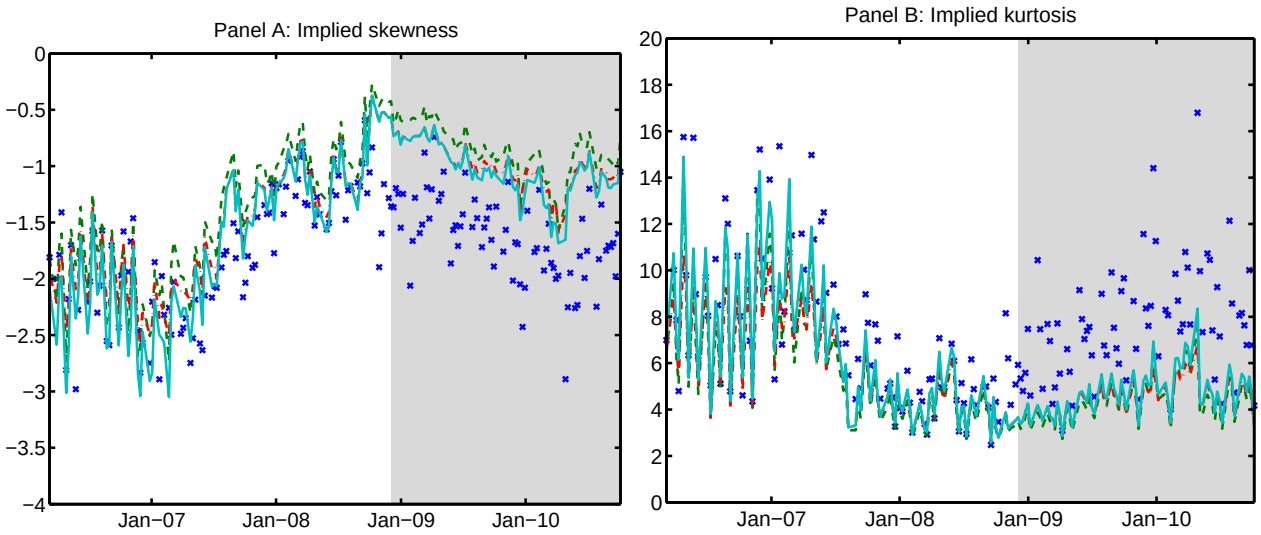


Figure 14: 1 month risk-neutral skewness and kurtosis of the distribution of the VIX implied by 1 month VIX options prices when estimating the SVJ2 (solid line), the SVJ (dashed line) and the SV2 (dashed dotted line) models over *Dataset 4* (indices as well as SPX and VIX options) from March 2006 to November 2008 (685 days). We use the method described in Bakshi, Kapadia, and Madan (2003). The shaded part of the graph represents the out-of-sample period, from December 2008 until October 2010.

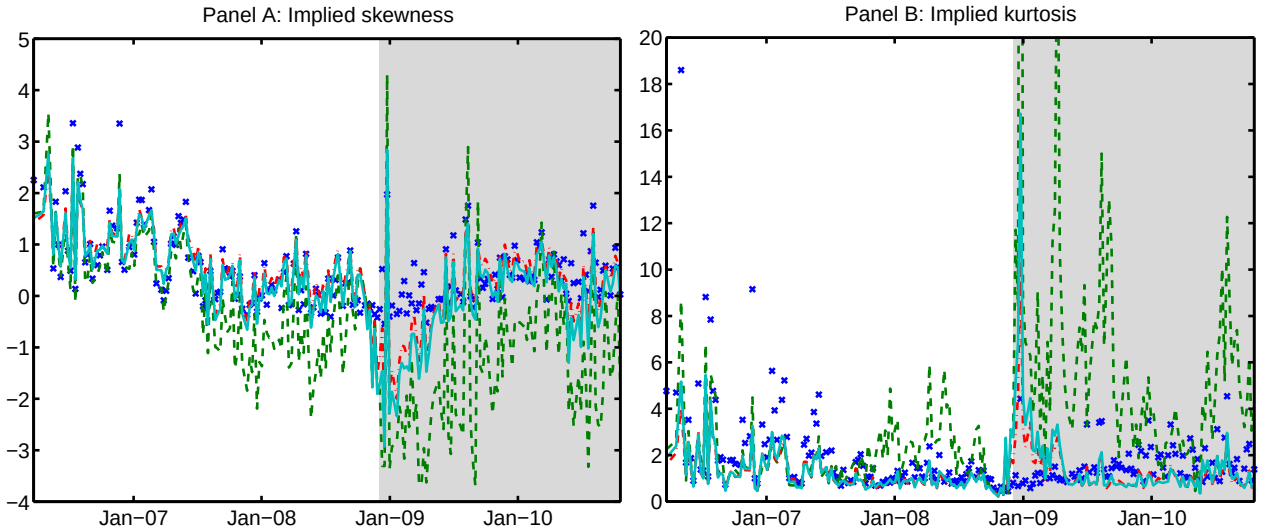
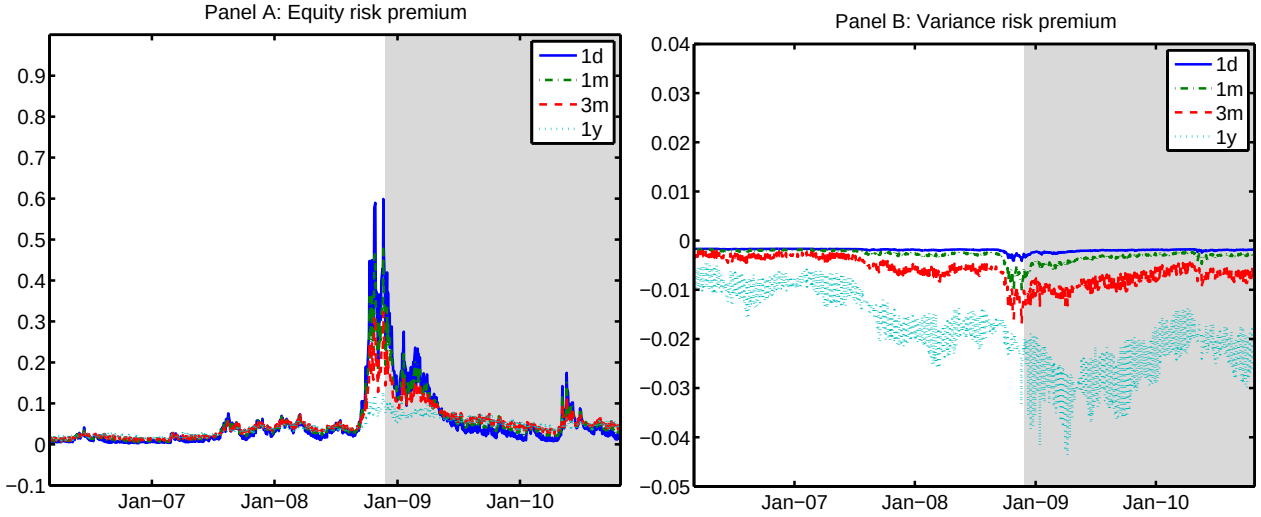


Figure 15: Equity and variance risk premia for different maturities when estimating the SVJ2 model using *Dataset 4* as estimation dataset from March 2006 to November 2008 (685 days). The shaded part of the graph represents the out-of-sample period, from December 2008 until October 2010.



ONLINE APPENDIX TO
Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets

This appendix provides the results of technical derivations.

1 Model specification under $\mathbb{P}$

Under the historical measure $\mathbb{P}$, the model is specified as follows:

$$dY_t = [-\lambda^{Y(\mathbb{P})}(v_{t^-}, m_{t^-})(\theta_Z^{(\mathbb{P})}(1,0,0) - 1) - \frac{1}{2}v_{t^-} + \gamma_t]dt + \sqrt{v_{t^-}}dW_t^{Y(\mathbb{P})} + dJ_t^{Y(\mathbb{P})} \tag{1.1}$$

$$dv_t = \kappa_v^{(\mathbb{P})}\left(\frac{\kappa_v^{(\mathbb{Q})}}{\kappa_v^{(\mathbb{P})}}m_{t^-} - v_{t^-}\right)dt + \sigma_v\sqrt{v_{t^-}}dW_t^{v(\mathbb{P})} + dJ_t^{v(\mathbb{P})} \tag{1.2}$$

$$dm_t = \kappa_m^{(\mathbb{P})}(\theta_m^{(\mathbb{P})} - m_{t^-})dt + \sigma_m\sqrt{m_{t^-}}dW_t^{m(\mathbb{P})} + dJ_t^{m(\mathbb{P})} \tag{1.3}$$

with:

$$\begin{aligned}
\gamma_t &= \eta_Y v_{t^-} + \lambda^Y(v_{t^-}, m_{t^-})(\theta_Z^{(\mathbb{P})}(1,0,0) - \theta_Z^{(\mathbb{Q})}(1,0,0)) \\
dW_t^{Y(\mathbb{P})} &= dW_t^{Y(\mathbb{Q})} - \eta_Y\sqrt{v_{t^-}}dt \\
dW_t^{v(\mathbb{P})} &= dW_t^{v(\mathbb{Q})} + \sqrt{v_{t^-}}\frac{\kappa_v^{(\mathbb{P})} - \kappa_v^{(\mathbb{Q})}}{\sigma_v}dt \\
dW_t^{m(\mathbb{P})} &= dW_t^{m(\mathbb{Q})} + \sqrt{m_{t^-}}\frac{\kappa_m^{(\mathbb{P})} - \kappa_m^{(\mathbb{Q})}}{\sigma_m}dt \\
\theta_m^{(\mathbb{P})} &= \frac{\kappa_m^{(\mathbb{Q})}\theta_m^{(\mathbb{Q})}}{\kappa_m^{(\mathbb{P})}}.
\end{aligned}$$

2 Coefficients of the VIX^2 formula

Table 11: Proposition 4.1 states that the VIX^2 depends on the instantaneous variance and level of mean reversion in an affine way. Here we give the values of coefficients playing a role in this proposition (see Appendix 1).

	A	B	G
$a_m \neq 0$ & $a_v \neq 0$ & $a_v \neq a_m$	$\frac{1}{a_v \tau_{\text{VIX}}}\left(e^{a_v \tau_{\text{VIX}}} - 1\right)$	$\frac{1}{\tau_{\text{VIX}}} \frac{h_v}{(a_m - a_v)} \left[\left(\frac{e^{a_m \tau_{\text{VIX}}} - 1}{a_m}\right) - \left(\frac{e^{a_v \tau_{\text{VIX}}} - 1}{a_v}\right)\right]$	$\frac{b_v}{a_v}\left[\left(\frac{e^{a_v \tau_{\text{VIX}}} - 1}{a_v \tau_{\text{VIX}}}\right) - 1\right] - b_m B$
$a_m \neq 0$ & $a_v \neq 0$ & $a_v = a_m$	$\frac{1}{a \tau_{\text{VIX}}}\left(e^{a \tau_{\text{VIX}}} - 1\right)$	$\frac{h_v}{a}\left(e^{a \tau_{\text{VIX}}} - \frac{1}{a \tau_{\text{VIX}}}\left(e^{a \tau_{\text{VIX}}} - 1\right)\right)$	$\frac{b_v}{a}\left[\left(\frac{e^{a \tau_{\text{VIX}}} - 1}{a \tau_{\text{VIX}}}\right) - 1\right] - b_m B$
$a_m \neq 0$ & $a_v = 0$	1	$\frac{h_v}{a_m}\left(\frac{1}{a_m \tau_{\text{VIX}}}\left(e^{a_m \tau_{\text{VIX}}} - 1\right) - 1\right)$	$\frac{1}{2} b_v \tau_{\text{VIX}} - b_m B$
$a_m = 0$ & $a_v \neq 0$	$\frac{1}{a_v \tau_{\text{VIX}}}\left(e^{a_v \tau_{\text{VIX}}} - 1\right)$	$\frac{h_v}{a_v}\left(\frac{1}{a_v \tau_{\text{VIX}}}\left(e^{a_v \tau_{\text{VIX}}} - 1\right) - 1\right)$	$\frac{c_m}{a_v}\left(B - \frac{1}{2} h_v \tau_{\text{VIX}}\right)$ $+ \frac{1}{a_v} \frac{\partial \theta_Z}{\partial \phi_v}(0,0,0) \lambda_0^v \left[\left(\frac{e^{a_v \tau_{\text{VIX}}} - 1}{a_v \tau_{\text{VIX}}}\right) - 1\right]$
$a_m = 0$ & $a_v = 0$	1	$\frac{1}{2} \tau_{\text{VIX}} h_v$	$\frac{1}{2} \tau_{\text{VIX}} \left[\frac{\partial \theta_Z}{\partial \phi_v}(0,0,0) \lambda_0^v + c_m h_v \frac{\tau_{\text{VIX}}}{3}\right]$

	$\hat{A}$	$\hat{B}$
$a_m \neq 0$	$\frac{e^{a_m \tau_{\text{VIX}}} - 1}{a_m \tau_{\text{VIX}}}$	$b_m\left(1 - \hat{A}\right)$
$a_m = 0$	1	$\frac{c_m \tau_{\text{VIX}}}{2}$

REFERENCES

Aït-Sahalia, Y., M. Karaman, and L. Mancini, 2012, "The Term Structure of Variance Swaps, Risk Premia and the Expectation Hypothesis," working paper, Swiss Finance Institute Working Paper.

Aït-Sahalia, Y., and R. Kimmel, 2007, "Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Financial Economics*, 83, 413–452.

Aït-Sahalia, Y., and A. Lo, 1998, "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *The Journal of Finance*, 53, 500ï¿½–547.

Alizadeh, S., M. Brandt, and M. Diebold, 2002, "Range-based estimation of stochastic volatility models," *The Journal of Finance*, 57, 1047–1091.

Andersen, T. G., L. Benzoni, and J. Lund, 2002, "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance*, 57, 1239–1284.

Andrews, D., 1991, "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817–858.

Bakshi, G., N. Kapadia, and D. Madan, 2003, "Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options," *The Review of Financial Studies*, 16(1), 101–143.

Bates, D. S., 1996, "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *Review of Financial Studies*, 9, 69–107.

——— , 2000, "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics*, 94, 181–238.

——— , 2003, "Empirical option pricing: a Retrospection," *Journal of Econometrics*, 116, 387 – 404.

——— , 2012, "U.S. stock market crash risk, 1926-2010," *Journal of Financial Economics*, 105(2), 229 – 259.

Bayer, C., J. Gatheral, and M. Karlsmark, 2013, "Fast Ninomiya-Victoir calibration of the Double-Mean-Reverting Model," working paper.

Bergomi, L., 2004, "Smile Dynamics," *Risk*, Sept., 117 – 123.

——— , 2005, "Smile Dynamics II," *Risk*, Oct., 67 – 73.

——— , 2008, "Smile Dynamics III," *Risk*, Oct., 90 – 96.

——— , 2009, "Smile Dynamics IV," *Risk*, Dec., 90 – 96.

Bollerslev, T., and V. Todorov, 2011, "Tails, Fears and Risk Premia," *Journal of Finance*, 66, 2165–2211.

Breeden, D., and R. H. Litzenberger, 1978, "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51, 621–651.

Broadie, M., M. Chernov, and M. Johannes, 2007, "Model Specification and Risk Premia: Evidence from Futures Options," *The Journal of Finance*, 62(3), 1453–1490.

Carr, P., and D. B. Madan, 1999, "Option valuation using the fast Fourier transform," *Journal of Computational Finance*, 2(4), 1–18.

Carr, P., and L. Wu, 2009, "Variance risk premiums," *Review of Financial Studies*, 22, 1311–1341.

CBOE, C. B. O. E., 2009, "THE CBOE VOLATILITY INDEX - VIX," working paper, Chicago Board Options Exchange, White paper.

Chernov, M., A. R. Gallant, E. Ghysels, and G. T. Tauchen, 2003, "Alternative Models of Stock Price Dynamics," *Journal of Econometrics*, 116, 225 – 257.

Chernov, M., and E. Ghysels, 2000, "A Study towards a Unified Approach to the Joint Estimation of Objective ad Risk Neutral Measures for the Purpose of Options Valuation," *Journal of Financial Economics*, 56, 407–458.

Christoffersen, P., and K. Jacobs, 2004, "The importance of the loss function in option valuation," *Journal of Financial Economics*, 72(2), 291 – 318.

Christoffersen, P., K. Jacobs, and K. Mimouni, 2010, "Models for S&P 500 Dynamics: Evidence from Realized Volatility, Daily Returns and Options Prices," *Review of Financial Studies*, 23(8), 3141–3189.

Chung, S., W. Tsai, Y. Wang, and P. Wenig, 2011, "The Information Content of the S&P 500 Index and VIX Options on the Dynamics of the S&P 500 Index," *The Journal of Futures Markets*, 31(12), 1170 – 1201.

Cont, R., and T. Kokholm, 2011, "A Consistent Pricing Model for Index Options and Volatility Derivatives," *Mathematical Finance*.

Demeterfi, K., E. Derman, M. Kamal, and J. Zou, 1999, "More than you ever wanted to know about volatility swaps," *Goldman Sachs Quantitative Strategies Research Notes*.

Detemple, J., and C. Osakwe, 2000, "The valuation of volatility options," *European Finance Review*, 4, 21–50.

Drimus, G. G., and E. W. Farkas, 2013, "Local Volatility of Volatility for the VIX Market," *Review of Derivatives Research*, (735), To appear.

Duan, J.-C., and C.-Y. Yeh, 2011, "Price and Volatility Dynamics Implied by the VIX Term Structure," working paper, NUS RMI Working Paper No. 11/05.

Duffie, D., J. Pan, and K. J. Singleton, 2000, "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.

Durham, G. B., 2012, "Risk-neutral modelling with affine and non-affine models," .

Egloff, D., M. Leippold, and L. Wu, 2010, "Valuation and Optimal Investing in Variance Swaps,"

Journal of Financial and Quantitative Analysis, 45(5), 1279–1310.

Eraker, B., 2004, "Do stock prices and volatility jump? Reconciling evidence from spot and option prices," *Journal of Finance*, 59, 1367–1403.

Fang, F., and C. Oosterlee, 2008, "A novel pricing method for European options based on Fourier-cosine series expansions," working paper, MPRA Paper 9319, University Library of Munich, Germany.

Ferriani, F., and S. Pastorello, 2012, "Estimating and Testing Non-Affine Option Pricing Models With a Large Unbalanced Panel of Options," *The Econometrics Journal*, 15(2), 171–203.

Gatheral, J., 2008, "Consistent Modeling of SPX and VIX options," The Fifth World Congress of the Bachelier Finance Society, Londo, July 2008.

Grünbichler, A., and F. Longstaff, 1996, "Valuing futures and options on volatility," *Journal of Banking and Finance*, 20, 985–1001.

Hansen, N., and A. Ostermeier, 1996, "Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation," *Proceedings of the 1996 IEEE Conference on Evolutionary Computation (ICEC 1996)*, pp. 312–317.

Jackwerth, J., 2000, "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies*, 13, 433–451.

Jiang, G. J., and Y. S. Tian, 2007, "Extracting Model-Free Volatility from Option Prices: An Examination of the VIX Index," *The Journal of Derivatives*, 14, 35–60.

Johannes, M. S., N. G. Polson, and J. R. Stroud, 2009, "Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices," *The Review of Financial Studies*, 22(7), 2759–2799.

Jones, C. S., 2003, "The Dynamics of Stochastic Volatility: Evidence From Underlying and Options Markets," *Journal of Econometrics*, 116, 181–224.

Kaeck, A., and C. Alexander, 2012, "Volatility dynamics for the S&P 500: Further evidence from non-affine, multi-factor jump diffusions," *Journal of Banking & Finance*, 36(11), 3110 – 3121.

Lian, G.-H., and S.-P. Zhu, 2011, "Pricing VIX options with stochastic volatility and random jumps," *Decisions in Economics and Finance*, pp. 1–18.

Lindström, E., J. Ströjby, M. Brodén, M. Wiktorsson, and J. Holst, 2008, "Sequential calibration of options," *Computational Statistics & Data Analysis*, 52(6), 2877–2891.

Mencía, J., and E. Sentana, 2013, "Valuation of VIX derivatives," *Journal of Financial Economics*, 108(2), 367 – 391.

Newey, W. K., and K. D. West, 1987, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Pan, J., 2002, "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series

Study," *Journal of Financial Economics*, 63, 3–50.

Papanicolaou, A., and R. Sircar, 2012, "A Regime-Switching Heston Model for VIX and S&P 500 Implied Volatilities," Submitted.

Pitt, M. K., and N. Shephard, 1999, "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94(446).

Rebonato, R., 2004, *Volatility and correlation*. Wiley, 2. ed. edn.

Rebonato, R., and T. Cardoso, 2004, "Unconstrained Fitting of Implied Volatility Surfaces Using a Mixture of Normals," *Journal of Risk*, 7(1), 55–74.

Sepp, A., 2008a, "Pricing Options on Realized Variance in the Heston Model with Jumps in Returns and Volatility," *Journal of Computational Finance*, 11(4), 33–70.

——— , 2008b, "VIX Option Pricing in a Jump-Diffusion Model," *Risk Magazine*, pp. 84–89.

Song, Z., and D. Xiu, 2012, "A Tale of Two Option Markets: State-Price Densities Implied from S&P 500 and VIX Option Prices," working paper, Chicago Booth Research Paper No 12-10 - Fama-Miller Working Paper.

Storn, R., 1996, "On the usage of differential evolution for function optimization," in *Biennial Conference of the North American Fuzzy Information Processing Society (NAFIPS)*, pp. 519–523.

Todorov, V., 2010, "Variance Risk Premium Dynamics: The Role of Jumps," *The Review of Financial Studies*, 23(1), 345–383.

Whaley, R. E., 1993, "Derivatives on Market Volatility: Hedging Tools Long Overdue," *The Journal of Derivatives*, 1, 71–84.

Wu, L., 2011, "Variance Dynamics: Joint Evidence from Options and High-Frequency Returns," *Journal of Econometrics*, 160, 280–287.

Quadratic Variance Swap Models

Damir Filipović, Elise Gourier and Loriano Mancini

I have and will be presenting this paper soon at:

- 7th World Congress of the Bachelier Finance Society, June 2012, Sydney, Australia
- 40th Annual Conference of the European Finance Association, August 2013, Cambridge, United Kingdom.

The most recent version of the paper is available at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2237512`.

Abstract

We introduce a novel class of term structure models for variance swaps. The multivariate state variable follows a diffusion process characterized by a quadratic diffusion function. The variance swap curve is quadratic in the state variable, and available in closed form in terms of a linear ordinary differential equation, greatly facilitating empirical analysis. Various goodness-of-fit tests show that quadratic models fit variance swaps on the S&P 500 remarkably well and outperform nested specifications, including popular affine models. An empirical study of a dynamic optimal portfolio in variance swaps and the S&P 500 reveals the versatility of quadratic models, and the economic value of variance swaps.

1 Introduction

A variance swap pays the difference between the realized variance of some underlying asset and the fixed variance swap rate. Variance swaps are actively traded at different maturities. This induces a term structure of variance swap rates, which reflects market expectations about future variance and provides important information for managing variance risk. Figure 16 shows variance swap rates on the S&P 500. The term structure takes a variety of shapes and exhibits rich dynamics. During low volatility periods, such as 2005–2006, the term structure is upward sloping. During financial crises, such as Fall 2008, the short-end spikes up, and the term structure becomes downward sloping. Having a model that captures such term structure movements appears to be crucial to consistently price variance swaps across different maturities or to optimally invest in such contracts. Surprisingly, the term structure of variance swap rates has received little attention in the literature.

We provide a novel class of flexible and tractable variance swap term structure models. The multivariate state variable follows a quadratic diffusion process characterized by linear drift and quadratic diffusion functions. Variance swap rates are quadratic in the state variable. The variance swap curve is available in closed form in terms of a linear ordinary differential equation, which greatly facilitates empirical applications. Higher order polynomial specifications are possible.

We perform an exhaustive specification analysis of the univariate quadratic model and of a parsimonious bivariate extension. Model identification is provided in terms of canonical representations. We also study univariate polynomial specifications of higher order. We fit these models to the term structure of variance swap rates on the S&P 500 shown in Figure 16, with five terms ranging from 2 months to 2 years, and daily quotes spanning from January 4, 1996 to June 7, 2010. Several statistical tests show that the bivariate quadratic model captures the term structure dynamics remarkably well. The quadratic state process is able to generate sudden large movements in the variance swap rates, and the quadratic variance swap model can produce a rich variety of term structure shapes, as observed empirically. Nested affine and other specifications are soundly rejected. We reach this conclusion using various likelihood-based tests (e.g., Giacomini and White (2006)), information theoretic criteria (i.e., Akaike and Bayesian Information Criteria), and Diebold–Mariano tests derived from variance swap pricing errors.

We find that the bivariate quadratic model produces better forecasts of variance swap rates than the univariate quadratic and polynomial models, as well as the martingale model. The latter uses today's variance swap rates as a prediction of future variance swap rates. Given the strong persistence of variance swap rates (first order autocorrelations are above 0.98), the martingale model is a challenging benchmark. When we regress future variance swap rates on model-based predictions of variance swap rates, we find that only the bivariate quadratic model has an intercept and a slope not statistically different from zero and one, respectively, and thus produces accurate forecasts. Moreover, only the bivariate model outperforms the martingale model, which in turn dominates the univariate quadratic and polynomial models. From an economic perspective, this suggests that the bivariate quadratic model captures ex-ante risk premiums embedded in variance swaps.

Equity and variance risk premiums induced by the bivariate quadratic model are economically sizable and exhibit significant time variation, which is in line with recent studies, e.g., Bollerslev and Todorov (2011), Aït-Sahalia, Karaman, and Mancini (2012), and Martin (2013). The equity risk premium is positive and countercyclical. The variance risk premium is mostly negative and procyclical. Overall, our empirical analysis suggests that the bivariate quadratic model offers a good trade-off between tractability and fitting accuracy of the term structure dynamics.

At least two features contribute to the popularity of variance swaps. First, hedging a variance swap is relatively easier than hedging other volatility derivatives. Indeed, the payoff of a variance swap can be replicated by dynamically trading in the underlying asset and a static position in a continuum of vanilla options with different strike prices and the same underlying and maturity date. In practice, of course, continuous trading is unfeasible and vanilla options exist only for a limited number of strike prices and may not exist at all for a given maturity date.[19] Second, the variance swap payoff is only sensitive to the realized variance over a desired and predetermined time horizon. Suppose an investor, who holds a broadly diversified portfolio, is concerned about volatility risk over the next month. Buying a variance swap on the S&P 500, with one month maturity, would provide a direct hedge against volatility risk. In contrast, taking positions on options and futures on the VIX index[20] would not provide an equally direct hedge.[21]

To assess the economic relevance of variance swaps, we study a dynamic optimal portfolio problem in variance swaps, a stock index and a risk free bond.[22] We solve for the optimal strategy of a power utility investor who maximizes the expected utility from terminal wealth. The variance swaps are on-the-run and rolled over at pre-specified arbitrary points in time. The optimal strategy, composed of the familiar myopic and intertemporal hedging terms (Merton (1971)), is derived in quasi-closed form. A Taylor series expansion of the intertemporal hedging term involves conditional moments of the state process, which are explicit in terms of a linear ordinary differential equation. We implement the optimal portfolio using actual 3-month and 2-year variance swap rates and S&P 500 returns. We find that the optimal portfolio weights in the variance swaps follow a short-long strategy, with a short position in the 2-year variance swap (to earn the negative variance risk premium), and a long position in the 3-month variance swap (to hedge volatility increases). This result is consistent with the empirical finding that long term variance swaps carry more variance risk premium and react less to volatility increases than short term variance swaps, e.g., Egloff, Leippold, and Wu (2010), and

[19]This led to a large literature analyzing and exploiting the various hedging errors when attempting to replicate a given variance swap, e.g., Neuberger (1994), Dupire (1993), Carr and Madan (1998), Demeterfi, Derman, Kamal, and Zou (1999), Britten-Jones and Neuberger (2000), Jiang and Tian (2005), Jiang and Oomen (2008), Carr and Wu (2009), and Carr and Lee (2010).

[20]The Chicago Board Options Exchange (CBOE) Market Volatility Index (VIX) is the 30-day variance swap rate on the S&P 500 quoted in volatility units. Carr and Wu (2006) provide an excellent history of the VIX index.

[21]It is so because the VIX index is the market expectation of the S&P 500 variance over the next 30 days. As time goes by, the VIX index, and derivatives on it, are sensitive to the S&P 500 variance expectation beyond the desired hedging horizon. In response to the need to trade volatility with more direct instruments, as pointed out by the CBOE, since December 2012 the CBOE has listed new contracts called "S&P 500 Variance Futures." These are exchange-traded, marked-to-market variance swaps on the S&P 500 with maturities ranging up to 2 years. See http://www.cfe.cboe.com/Products/Spec_VA.aspx.

[22]Egloff, Leippold, and Wu (2010) study a similar investment problem. However, there are several differences between the two studies, which are discussed in detail at the end of Section 6.1.

Aït-Sahalia, Karaman, and Mancini (2012). We also find that optimal weights in variance swaps show strong periodic patterns, which depend on the maturity and roll-over date of the contracts, and which are mainly borne by the intertemporal hedging demand. The optimal weight in the stock index is positive (to earn the equity risk premium).

We consider two relative risk aversion levels, 5 and 1. The first is an average value in survey data.[23] The second corresponds to logarithmic utility. Optimal portfolio weights for both levels share the patterns described above. However, the respective wealth trajectories are largely different. The more risk averse investor takes on smaller positions than the log-investor, in absolute value. This results in a smooth and steady growth of her wealth over time, which is largely unaffected by market declines. In contrast, the wealth trajectory of the log-investor exhibits large fluctuations, even more than the S&P 500. This suggests that variance swaps can be used either to achieve stable wealth growth or to seek additional risk premiums, depending on the risk profile of the investor. Rebalancing the portfolio less frequently than daily, such as monthly and yearly, leads to similar results.

To further understand the performance of optimal portfolios under different economic scenarios and in terms of expected utility, we run a Monte Carlo simulation. We compare optimal portfolios in variance swaps, stock index, and bond to the stock index, and for the log-investor, to the optimal portfolio in stock index and bond. We find that the optimal portfolio including variance swaps significantly outperforms the others, both in terms of certainty equivalent and Sharpe ratio. This suggests that variance swaps have a significant economic value for risk averse investors.

Our paper is related to various strands of the literature. A fast growing literature studies the variance risk premium and its impact on asset prices, e.g., Jiang and Tian (2005), Carr and Wu (2009), Bollerslev, Tauchen, and Zhou (2009), Todorov (2010), Bollerslev and Todorov (2011), Drechsler and Yaron (2011), and Mueller, Vedolin, and Yen (2011). This line of research focuses almost exclusively on a single maturity. As mentioned above, the term structure of variance swap rates has remained unexplored until recently, e.g., Amengual (2009), Egloff, Leippold, and Wu (2010), and Aït-Sahalia, Karaman, and Mancini (2012).[24] Part of the reason could be that variance swap data became available only recently. We contribute to this line of research by proposing a novel quadratic term structure model, assessing its empirical performance, and studying dynamic optimal portfolios in this setting.

There is an extensive literature on term structure models for interest rates. This literature mainly focuses on affine term structure models, where the zero-coupon yield curve is affine in the state variable which follows an affine diffusion process.[25] The loadings in turn are given in terms of a non-linear ordinary differential equation.[26] Quadratic and higher order polynomial specifications of the yield

[23] Most of the survey data suggests values of the relative risk aversion between 0.23 and 8, e.g., Meyer and Meyer (2005).
[24] Fusari and Gonzalez-Perez (2012) provide a related study based on variance swap rates computed using vanilla options on the S&P 500 index. Buehler (2006) and Gatheral (2008) mainly focus on theoretical models for the term structure of variance swap rates.
[25] Affine diffusion processes are nested in our class of quadratic diffusion processes.
[26] See, e.g., Duffie and Kan (1996), Dai and Singleton (2000), Duffie, Pan, and Singleton (2000), Duffie, Filipović, and Schachermayer (2003), and Collin-Dufresne, Goldstein, and Jones (2008). Dai and Singleton (2003), and Duarte (2004) discuss some limitations of affine term structure models. Various extensions of affine models have been suggested by, e.g., Constantinides (1992), Goldstein (2000), Leippold and Wu (2002), Collin-Dufresne and Goldstein (2002), Ahn, Dittmar, and Gallant (2002), Kimmel (2004), and Collin-Dufresne, Goldstein, and Jones (2009).

curve are limited if not inexistant, Filipović (2002), and Chen, Filipović, and Poor (2004). These limitations do not exist for the variance swap curve. This allows us to define the class of generic quadratic variance swap models, where the spot variance is a quadratic, or higher order polynomial, function of the state variable which follows a quadratic diffusion process. The resulting variance swap curve is quadratic, or higher order polynomial, in the state variable, and the loadings are given in terms of a linear ordinary differential equation.

Several papers have studied dynamic optimal portfolios with stochastic investment opportunity set, e.g., Kim and Omberg (1996), Brennan and Xia (2002), Chacko and Viceira (2005), Sangvinatsos and Wachter (2005), and Liu (2007). These papers mainly focus on optimal investment in a stock and a bond. Liu and Pan (2003) extend the investment opportunity set to options, and Egloff, Leippold, and Wu (2010) to variance swaps in an affine setting. We also study optimal portfolios including variance swaps. As discussed in Section 6, our quadratic setting yields optimal strategies that are significantly different from the ones in Egloff, Leippold, and Wu (2010).

The structure of the paper is as follows. Section 2 presents variance swaps. Section 3 introduces quadratic variance swap models. Section 4 discusses model estimates. Section 5 studies optimal portfolios in variance swaps, stock index and risk free bond. Section 6 investigates the empirical performance of optimal portfolios. Section 7 concludes. Technical derivations and proofs are collected in an online appendix.

2 Variance Swaps

We fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ where $\mathbb{P}$ is the objective probability measure. Let S_t be a continuous semimartingale modeling the price process of a stock index with spot variance process v_t. Let $\mathbb{Q}$ be an equivalent risk neutral measure under which the risk free discounted price process follows a local martingale.

Let $t = t_0 < t_1 < \cdots < t_n = T$ denote the trading days over a given time period $[t, T]$. The annualized realized variance is the annualized sum of squared log-returns over the given time horizon:

$$\mathrm{RV}(t,T) = \frac{252}{n} \sum_{i=1}^{n} \left(\log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2 .$$

It is known that, as $\sup_{i=1,\ldots,n} (t_i - t_{i-1}) \to 0$, the realized variance converges in probability to the quadratic variation of the log-price:

$$\sum_{i=1}^{n} \left(\log \frac{S_{t_i}}{S_{t_{i-1}}} \right)^2 \xrightarrow{\mathbb{P}} \int_t^T v_s \, ds.$$

This approximation is commonly adopted in practice (e.g., Egloff, Leippold, and Wu (2010)) and quite accurate at a daily sampling frequency (e.g., Broadie and Jain (2008), and Jarrow, Kchia, Larsson,

and Protter (2013)), as is the case in our dataset.[27]

A variance swap initiated at t with maturity T, or term $T-t$, pays the difference between the annualized realized variance $\mathrm{RV}(t,T)$ and the variance swap rate $\mathrm{VS}(t,T)$ fixed at t.[28] By convention, the variance swap rate is such that the variance swap contract has zero value at inception. No arbitrage implies that

$$\mathrm{VS}(t,T) = \frac{1}{T-t}\mathbb{E}_{\mathbb{Q}}\left[\int_t^T v_s\,ds \mid \mathcal{F}_t\right] \tag{2.1}$$

where $\mathbb{E}_{\mathbb{Q}}$ denotes the expectation under $\mathbb{Q}$, and we assume that the risk free rate and the spot variance are independent processes under $\mathbb{Q}$.

To consistently price variance swaps and capture the term structure of volatility risk, it is crucial to design models for the entire variance swap curve $T \mapsto \mathrm{VS}(t,T)$. In view of (2.1), this boils down to modeling the spot variance process $v_t = \mathrm{VS}(t,t)$ under $\mathbb{Q}$. These models should be analytically tractable and yet flexible enough to reproduce the empirical features of variance swap rates. Any positive continuous semimartingale whose spot variance process coincides with v_t is then a consistent price process in the sense that $\mathrm{VS}(t,T)$ is the corresponding variance swap rate.

Our approach easily extends to semimartingale price processes with jumps. The spot variance is then to be set to $v_t = \sigma_t^2 + \int_{\mathbb{R}} x^2\,\nu_t(x)\,dx$ where σ_t^2 denotes the spot variance of the continuous martingale part and $\nu_t(x)$ is the $\mathbb{Q}$-compensator of the jumps of the log-price.[29]

It is instructive to draw an analogy between the term structure of variance swap and interest rates. The variance swap curve reflects market expectations about future changes in spot variance, (2.1). The financial variable in interest rate models corresponding to the spot variance v_t is the risk free short rate r_t. Market expectations about future changes in short rates are expressed in terms of the zero-coupon yield curve

$$y(t,T) = -\frac{1}{T-t}\log \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-\int_t^T r_s\,ds} \mid \mathcal{F}_t\right],$$

with short-end given by $y(t,t) = r_t$. Clearly, the yield curve is a non-linear function of the short rate process. In contrast, the variance swap curve is a linear function of the spot variance process. This linear relationship gives greater flexibility for the specification of analytically tractable term structure models for variance swap than for interest rates. Indeed, most common factor models for the term structure of interest rates are affine term structure models. The short rate is specified as an affine function of the state variable which follows an affine diffusion process. The resulting yield curve is affine in the state variable, and the loadings are given as solutions to a non-linear ordinary

[27] Market microstructure noise, while generally a concern in high frequency inference, is largely a non-issue at the level of daily returns.

[28] As the difference is in variance units, the payoff is converted in dollar units via a suitable notional amount.

[29] Note that the spot variance $v_t = \sigma_t^2 + \int_{\mathbb{R}} x^2\,\nu_t(x)\,dx$ and the corresponding variance swap rates $\mathrm{VS}(t,T)$ can be continuous processes even if the underlying log-price exhibits jumps, i.e., $\nu_t(x)$ is non-zero. Aït-Sahalia, Karaman, and Mancini (2012) provide empirical evidence that variance swap rates on the S&P 500 contain a non-zero $\nu_t(x)$, and model v_t as an affine diffusion. For the empirical analysis in this paper we model variance swap rates using a general quadratic diffusion without specifying the stock index dynamics. For the optimal portfolio problem in Section 5, for tractability, we specify the stock index as a continuous process.

differential equation, e.g., Duffie and Kan (1996), and Dai and Singleton (2000). Specifying the short rate as a quadratic function of the state variable is possible. But it generically requires that the state variable follows a Gaussian process, e.g., Ahn, Dittmar, and Gallant (2002), Chen, Filipović, and Poor (2004), and Liu (2007).[30] Moreover, there exists no consistent polynomial specification of the yield curve beyond second order, Filipović (2002). These limitations do not exist for variance swap term structure models, and this flexibility is exploited here.

3 Quadratic Variance Swap Models

Let X_t be a diffusion process in some state space $\mathcal{X} \subset \mathbb{R}^m$, solving the stochastic differential equation (SDE)

$$dX_t = \mu(X_t)\,dt + \Sigma(X_t)\,dW_t \tag{3.1}$$

where W_t is a standard d-dimensional Brownian motion under the risk neutral measure $\mathbb{Q}$, and $\mu(x)$ and $\Sigma(x)$ are $\mathbb{R}^m$- and $\mathbb{R}^{m\times d}$-valued functions on $\mathcal{X}$, for some integers $m, d \geq 1$. The process X_t has the following quadratic structure:

Definition 3.1. *The diffusion X_t is called* quadratic *if its drift and diffusion functions are linear and quadratic in the state variable:*

$$\mu(x) = b + \beta\, x \tag{3.2}$$

$$\Sigma(x)\Sigma(x)^\top = a + \sum_{k=1}^{m} \alpha^k x_k + \sum_{k,l=1}^{m} A^{kl} x_k x_l \tag{3.3}$$

for some parameters $b \in \mathbb{R}^m$, $\beta \in \mathbb{R}^{m\times m}$, and a, α^k, $A^{kl} \in \mathbb{S}^m$ with $A^{kl} = A^{lk}$, where $\mathbb{S}^m$ denotes the set of symmetric $m \times m$-matrices, and $^\top$ denotes transposition.

An m-factor *quadratic variance swap model* is obtained by imposing that the spot variance is a quadratic function of the state variable:

$$v_t = g(X_t) \tag{3.4}$$

with $g(x) = \phi + \psi^\top x + x^\top \pi\, x$, for some parameters $\phi \in \mathbb{R}$, $\psi \in \mathbb{R}^m$, and $\pi \in \mathbb{S}^m$. The following proposition justifies the terminology of quadratic variance swap model.

Proposition 3.1. *Under the above assumptions, the quadratic variance swap model admits a* quadratic term structure. *That is, the variance swap rates are quadratic in the state variable:*

$$\mathrm{VS}(t,T) = \frac{1}{T-t} G(T-t, X_t) \tag{3.5}$$

with $G(\tau, x) = \Phi(\tau) + \Psi(\tau)^\top x + x^\top \Pi(\tau)\, x$, where the functions $\Phi : [0,+\infty) \to \mathbb{R}$, $\Psi : [0,+\infty) \to \mathbb{R}^m$,

[30] Liu (2007) considers mixtures of quadratic-Gaussian and affine components in a specific setup.

and $\Pi : [0, +\infty) \to \mathbb{S}^m$ *satisfy the linear ordinary differential equations*

$$\begin{aligned}
\frac{d\Phi(\tau)}{d\tau} &= \phi + b^\top \Psi(\tau) + \operatorname{tr}(a\,\Pi(\tau)), && \Phi(0) = 0 \\
\frac{d\Psi(\tau)}{d\tau} &= \psi + \beta^\top \Psi(\tau) + 2\Pi(\tau)b + \alpha \cdot \Pi(\tau), && \Psi(0) = 0 \\
\frac{d\Pi(\tau)}{d\tau} &= \pi + \beta^\top \Pi(\tau) + \Pi(\tau)\beta + A \cdot \Pi(\tau), && \Pi(0) = 0
\end{aligned} \tag{3.6}$$

where we define the tensor operations $(\alpha \cdot \Pi)_k = \operatorname{tr}(\alpha^k\,\Pi)$ *and* $(A \cdot \Pi)_{kl} = \operatorname{tr}(A^{kl}\,\Pi)$.

Proof. The assertion follows from (2.1) and Lemma 1.1 in Appendix 1 with $f(\tau, x) = \partial G(\tau, x)/\partial\tau$. □

Appendix 2 shows that, under mild technical conditions, the converse to Proposition 3.1 also holds true: a quadratic term structure implies that the spot variance function and the state diffusion process X_t be necessarily quadratic. This result implies that our quadratic model framework is exhaustive as we do not miss any other diffusion specification which is consistent with a quadratic term structure.

We also specify an $\mathbb{R}^d$-valued process for the market price of risk, Λ, such that $dW_t^{\mathbb{P}} = dW_t - \Lambda_t\,dt$ is a $\mathbb{P}$-Brownian motion and the identity $\Sigma(X_t)\,\Lambda_t = \Upsilon_0 + \Upsilon_1 X_t$ holds for some parameters $\Upsilon_0 \in \mathbb{R}^m$ and $\Upsilon_1 \in \mathbb{R}^{m\times m}$. This implies that the $\mathbb{P}$-dynamics of X_t are of the form

$$dX_t = (b + \Upsilon_0 + (\beta + \Upsilon_1)X_t)\,dt + \Sigma(X_t)\,dW_t^{\mathbb{P}}.$$

Thus, the process X_t follows a quadratic diffusion under $\mathbb{P}$ as well. The properties of X_t derived from the quadratic structure hold under $\mathbb{Q}$ as well as under $\mathbb{P}$.

It follows by inspection that an affine transformation of the state, $X_t \mapsto c + \gamma\,X_t$, preserves the quadratic property (3.2)–(3.3) of X_t and the quadratic term structure (3.5). From an econometric viewpoint, this implies that the above general model is not identifiable. This calls for a canonical representation. A full specification analysis of general multi-factor quadratic models is beyond the scope of this paper.[31] In the following sections, we first provide an exhaustive specification analysis for the univariate quadratic model. We then study a bivariate extension and univariate polynomial specifications of higher order. Model identification is asserted in terms of canonical representations.

[31] This would require to find necessary and sufficient conditions on the model parameters and the state space $\mathcal{X}$ such that the multivariate quadratic diffusion X_t be well-defined in $\mathcal{X}$. The matrix-valued quadratic form on the right hand side of (3.3) needs to be positive semi-definite for all $x \in \mathcal{X}$. Moreover, it has to vanish in the direction orthogonal to the boundary at all boundary points, in order that the state space be invariant under the dynamics of X_t. Hence the state space $\mathcal{X}$ is specified by the zeros of quadratic forms on $\mathbb{R}^m$. The zero level sets of quadratic forms on $\mathbb{R}^m$ are complex geometric objects, and the canonical classification of quadratic diffusions would at least require an exhaustive classification of such zero level sets.

3.1 Univariate Quadratic Model

In this section, let $m = d = 1$ and consider a univariate quadratic diffusion

$$dX_t = (b + \beta X_t)\, dt + \sqrt{a + \alpha X_t + A X_t^2}\, dW_t \tag{3.7}$$

on some interval $\mathcal{X}$ in $\mathbb{R}$ and for some real parameters b, β, a, α, and $A \geq 0$. The linear ordinary differential equations (3.6) simplify to (3.1) in Appendix 3.

The invariance of quadratic processes with respect to affine transformations allows us to distinguish exactly three equivalence classes of quadratic processes on unbounded intervals with a canonical representation each. In other words, any univariate quadratic process (on unbounded intervals and possibly after an affine transformation) necessarily falls in one of the three equivalence classes. The three canonical representations are identifiable, and thus can be estimated using variance swap data. The proof is given in Appendix 4.

Proposition 3.2. *Denote the discriminant of the diffusion function of X_t by $D = \alpha^2 - 4Aa$. The quadratic process X_t falls in one of the following three equivalence classes:*

- *Class 1: either $A > 0$ and $D < 0$, or $A = \alpha = 0$ and $a > 0$. The canonical representation is specified by $\mathcal{X} = \mathbb{R}$, $b \geq 0$, $\beta \in \mathbb{R}$, $a = 1$, $\alpha = 0$, $A \geq 0$, and hence*

  $$dX_t = (b + \beta X_t)\, dt + \sqrt{1 + A X_t^2}\, dW_t.$$

  Note that for $A = 0$ we obtain a Gaussian process.

- *Class 2: either $A > 0$ and $D = 0$, or $A = \alpha = a = 0$. The canonical representation is specified by $\mathcal{X} = (0, +\infty)$, $b = 1$ or 0, $\beta \in \mathbb{R}$, $a = 0$, $\alpha = 0$, $A \geq 0$, and hence*

  $$dX_t = (b + \beta X_t)\, dt + \sqrt{A X_t^2}\, dW_t.$$

  Note that for $A = 0$ we obtain a deterministic process.

- *Class 3: either $A > 0$ and $D > 0$, or $A = 0$ and $\alpha \neq 0$. The canonical representation is specified by $\mathcal{X} = [0, +\infty)$, $b \geq 0$, $\beta \in \mathbb{R}$, $a = 0$, $\alpha = 1$, $A \geq 0$, and hence*

  $$dX_t = (b + \beta X_t)\, dt + \sqrt{X_t + A X_t^2}\, dW_t.$$

  The boundary point 0 is not attainable if and only if $b \geq 1/2$, in which case we can choose $\mathcal{X} = (0, +\infty)$. Note that for $A = 0$ we obtain an affine process.

Remark 3.1. *For $A < 0$ and $D > 0$, the state space $\mathcal{X}$ becomes bounded. The canonical representation for this equivalence class is the Jacobi process on $\mathcal{X} = [0, 1]$. We do not consider this case, as here we focus on state processes on unbounded state spaces.*

3.2 Bivariate Quadratic Model

In this section, we consider a bivariate extension of the above univariate quadratic variance swap model. Higher dimensional extensions are conceptually straightforward, but these models would be quite difficult to estimate because of the large number of parameters. Our empirical analysis below shows that a bivariate model provides a good fit to variance swap data, thus higher dimensional extensions do not appear to be practically relevant.

Let $m = 2$ and consider a bivariate quadratic diffusion $X_t = (X_{1t}, X_{2t})^\top$ of the form

$$\begin{aligned} dX_{1t} &= (b_1 + \beta_{11}\, X_{1t} + \beta_{12}\, X_{2t})\, dt + \sqrt{a_1 + \alpha_1 X_{1t} + A_1 X_{1t}^2}\, dW_{1t} \\ dX_{2t} &= (b_2 + \beta_{22}\, X_{2t})\, dt + \sqrt{a_2 + \alpha_2 X_{2t} + A_2 X_{2t}^2}\, dW_{2t} \end{aligned}$$

with $\beta_{12} \geq 0$ and $X_{2t} \geq 0$. The components X_{1t} and X_{2t} are instantaneously uncorrelated and only interact via the drift term. The spot variance function is assumed to depend on X_{1t} only,

$$g(x) = \phi + \psi x_1 + \pi x_1^2,$$

where $x = (x_1, x_2)$, for some real parameters ϕ, ψ and π. Hence X_{1t} drives the spot variance, while X_{2t} determines the stochastic mean reversion level, $-(b_1 + \beta_{12} X_{2t})/\beta_{11}$, of X_{1t}. The linear ordinary differential equations (3.6) simplify to (3.2) in Appendix 3.

The admissible specifications for X_{2t} are either Class 2 or 3 with the corresponding canonical representations given by Proposition 3.2. The diffusion function of X_{1t} can be of any Class 1–3 with the corresponding canonical representations from Proposition 3.2. Imposing $b_1 = 0$ when the diffusion function of X_{1t} is in Class 1 or 2, and $b_1 = 0$ or $1/2$ when it is in Class 3, ensures that the bivariate quadratic model is identified. This is proved in Appendix 5. The univariate quadratic model is nested in the bivariate model, setting X_{2t} to a positive constant value.

To keep the model parsimonious, a risk premium is attached only to the first Brownian motion, W_{1t}. The market price of risk process is then

$$\Lambda_t = \left(\frac{\lambda_0 + \lambda_1 X_{1t}}{\sqrt{a_1 + \alpha_1 X_{1t} + A_1 X_{1t}^2}}, 0 \right)^\top. \tag{3.8}$$

The parameter λ_0 may take any real value if the diffusion function of X_{1t} is in Class 1, $\lambda_0 \geq 0$ if the diffusion function of X_{1t} is in Class 2 or in Class 3 along with $b_1 = 1/2$, and $\lambda_0 = 0$ otherwise. It follows from Cheridito, Filipović, and Kimmel (2007) that the change of measure $\mathbb{P} \sim \mathbb{Q}$ is well defined under these conditions.

3.3 Univariate Polynomial Model

An important property of quadratic diffusion processes is that their conditional nth moments are available in closed form as polynomials of degree n in the state variables. This is in fact the reason why in Proposition 3.1 we obtain the closed form quadratic expression for $G(T-t, X_t)$. Indeed, $\partial G(T-t, X)/\partial T$ is simply the $\mathcal{F}_t$-conditional moment of the quadratic polynomial $g(X_T)$ in X_T. This polynomial preserving property of X_t suggests a natural extension of the quadratic variance swap models, namely higher order polynomial variance swap models. Here we discuss the univariate case. The multivariate case is a straightforward but notationally cumbersome extension.

As in Section 3.1, we consider the univariate quadratic diffusion process (3.7). The following proposition formalizes the polynomial preserving property of X_t. The proof is given in Appendix 6.

Proposition 3.3. *The $(N+1)$-row vector of the first N $\mathcal{F}_t$-conditional moments of $X_{t+\tau}$ with $\tau \geq 0$ is given by*

$$\left(1,\, \mathbb{E}\left[X_{t+\tau} \mid \mathcal{F}_t\right], \cdots, \mathbb{E}\left[X_{t+\tau}^N \mid \mathcal{F}_t\right]\right) = \left(1,\, X_t,\, \cdots, X_t^N\right) \mathrm{e}^{B\,\tau}$$

where B is an upper triangular $(N+1)\times(N+1)$ matrix defined in (6.2) in Appendix 6, and $\mathrm{e}^{B\,\tau}$ denotes the matrix exponential of $B\tau$.

A *polynomial variance swap model* is then obtained by specifying the spot variance as a polynomial function of the state variable, $v_t = p_0 + p_1 X_t + \cdots + p_N X_t^N$, for some parameters $p_i \in \mathbb{R}, i = 0, \ldots, N$. The following corollary is an immediate consequence of Proposition 3.3.

Corollary 3.1. *Under the above assumptions, the polynomial variance swap model admits a* polynomial term structure. *That is, the variance swap rates are polynomial of degree N in X_t:*

$$\mathrm{VS}(t,T) = \frac{1}{T-t}\left(P_0(T-t) + P_1(T-t)X_t + \cdots + P_N(T-t)X_t^N\right) \tag{3.9}$$

where the functions $P_i : [0, +\infty) \to \mathbb{R}$ satisfy the linear ordinary differential equations

$$\frac{dP(\tau)}{d\tau} = p + B\,P(\tau), \quad P(0) = 0 \tag{3.10}$$

where $P(\tau) = (P_0(\tau),\, P_1(\tau), \ldots, P_N(\tau))^\top$ and $p = (p_0,\, p_1, \ldots, p_N)^\top$.

It follows by inspection that the system (3.10) is equivalent to (3.6) for $N = 2$, with loadings $\Phi(\tau) = P_0(\tau)$, $\Psi(\tau) = P_1(\tau)$, and $\Pi(\tau) = P_2(\tau)$.

4 Model Estimation

In this section, we fit the variance swap models in Sections 3.1–3.3 directly to variance swap rates on the S&P 500, without specifying the index dynamic. An advantage of this approach is that model estimates are not impaired by potential misspecifications of the index dynamic and allows for

a thorough comparison of the variance swap models.

4.1 Dataset

Our dataset consists of daily closing over-the-counter quotes of variance swap rates on the S&P 500 index, with fixed terms at 2, 3, and 6 months, and 1 and 2 years.[32] It spans from January 4, 1996 to June 7, 2010, and includes 3,626 observations for each term. Standard statistical tests do not detect any day-of-the-week-effect, so we use all available daily data. An interesting feature of this dataset is that terms, rather than maturities, are fixed. This facilitates the comparison of the term structure over time, without using any interpolation method to recover variance swap rates for a specific term.

Figure 16 shows the term structure of variance swap rates over time and suggests that variance swap rates are mean-reverting, volatile, with spikes and clustering during the major financial crises over the last 15 years, and historically high values during the financial crisis in Fall 2008. While most term structures are upward sloping (48% of our sample), they can also be $\cup$-shaped (23% of our sample) and more rarely downward sloping or $\cap$-shaped.[33] The bottom and peak of the $\cup$- and $\cap$-shaped parts of the term structures, can be anywhere at the 3 or 6 months or 1 year term. The slope of the term structure, measured as the difference between the 2-year and 2-month variance swap rates, shows a strong negative relation to the contemporaneous level of volatility. Thus, in high volatility periods, the short-end of the term structure (variance swap rates with 2 or 3 months term) rises more than the long-end, producing downward sloping term structures.

Table 12 provides summary statistics of our dataset. We split the sample in two parts. The first part ranges from January 4, 1996 to April 2, 2007, includes 2,832 daily observations (about 3/4 of the whole sample), and will be used for in-sample analysis and model estimation. The second part ranges from April 3, 2007 to June 7, 2010, includes 794 daily observations, and will be used for out-of-sample analysis, including model validation. The out-of-sample analysis appears to be particularly interesting as the sample period covers the recent financial crisis, a period of unprecedented market turmoil, which was not experienced in the in-sample period.

For the sake of interpretability, we follow market practice and report variance swap rates in volatility percentage units, i.e., $\sqrt{\mathrm{VS}(t,T)} \times 100$. Various empirical regularities emerge from Table 12. The mean level of variance swap rates is slightly but strictly increasing with term. The standard deviation, skewness and kurtosis of variance swap rates are decreasing with term. Unreported first order autocorrelations of variance swap rates range between 0.984 and 0.995, are slightly increasing with the term and imply a mean half-life of shocks between 43 and 138 days.[34] This confirms that mean reversion is present in the time series and suggests that long term variance swap rates are more persistent than short term rates. Comparing in- and out-of-sample statistics reveals a significant increase in level and

[32] We thank Mika Kastenholz from Credit Suisse for providing us with the variance swap data.

[33] On some occasions, the term structure is $\sim$-shaped, but the difference between, for e.g., the 2 and 3 months variance swap rates is virtually zero and this term structure is nearly $\cup$-shaped.

[34] The half-life H is defined as the time necessary to halve a unit shock and solves $\varrho^H = 0.5$, where ϱ is the first order autocorrelation coefficient.

volatility of variance swap rates, mainly due to the market turmoil in Fall 2008.

A Principal Component Analysis (PCA) shows that the first principal component explains about 95.3% of the total variance of variance swap rates and can be interpreted as a level factor, while the second principal component explains an additional 3.8% and can be interpreted as a slope factor.[35] This finding is somehow expected because PCA of several other term structures, such as bond yields, produces qualitatively similar results. Less expected is that two factors explain nearly all the variance of variance swap rates, i.e., 99.1%. Repeating the PCA for various subsamples produces little variation in the first two factors and explained total variance.

Table 12 also shows summary statistics of ex-post realized variance of S&P 500 returns for various terms. All statistics of realized variances share qualitatively the same features as those of the variance swap rates. The main difference is that, especially during the in-sample period, realized variances tend to be lower and more volatile, positively skewed and leptokurtic than variance swap rates. This difference highlights the profitability and riskiness of shorting variance swaps, earning large negative variance risk premiums embedded in such contracts. The ex-post variance risk premium is defined as the average realized variance minus the variance swap rate, which is simply the average payoff of a long position in the respective variance swap. The corresponding summary statistics are reported in the last panel of Table 12. In the in-sample period, ex-post variance risk premiums are negative and, except for the longest maturity, increasing in absolute value with the term. Notably, ex-post Sharpe ratios from shorting variance swaps also increase with their term, ranging from 0.60 (= 1.67/2.80) for 2-month variance swaps to 0.85 (= 2.15/2.54) for 1-year variance swaps. This suggests that it is more profitable on average to sell long term than short term variance swaps. In the out-of-sample period, the opposite holds as short term variance swap rates increase proportionally more than long term variance swap rates, making it more profitable, ex-post, to buy long term variance swaps.

To summarize, the term structure of variance swap rates exhibits rich dynamics, challenging any term structure model. Whether our quadratic models are flexible enough to fit variance swap rates is an empirical question that we address in the following two sections.

4.2 Model Estimates

The state process X_t driving the term structure is not observed. We use the extended Kalman filter to extract the latent state and compute the likelihood of a particular model. Duffee and Stanton (2004), among others, provide a detailed description of the method. Here we briefly discuss the implementation of the filter.

Let $\mathbf{VS}(t)$ denote the five-dimensional vector of variance swap rates with terms τ_j equal to 2, 3, 6 months, and 1 and 2 years observed at time t. Define the vector-valued function $\mathbf{H}(x)$ with j-th component given by $G(\tau_j, x)/\tau_j$, see (3.5), and denote by $D_x\mathbf{H}(x)$ its derivative. The measurement

[35] To save space, factor loadings are not reported, but are available from the authors upon request.

equation is then linearized as follows:

$$\mathbf{VS}(t_i) = \mathbf{H}(\hat{X}_{t_i|t_{i-1}}) + D_x\mathbf{H}(\hat{X}_{t_i|t_{i-1}})\,(X_{t_i} - \hat{X}_{t_i|t_{i-1}}) + \eta_{t_i} \tag{4.1}$$

where $\hat{X}_{t_i|t_{i-1}}$ denotes the time-t_{i-1} prediction of X_{t_i}, η_{t_i} is a normal zero-mean error term, and $t_i - t_{i-1} \equiv 1/252$ is one day.

The state transition equation in (3.1) is discretized using an Euler scheme at daily frequency and parameter estimates are obtained by maximizing the (quasi) log-likelihood function

$$\sum_{i=1}^{\mathcal{N}} -\frac{1}{2}\left[5\log(2\pi) + \log|V_{t_i|t_{i-1}}| + e_{t_i}^{\top}V_{t_i|t_{i-1}}^{-1}e_{t_i}\right] \tag{4.2}$$

where $e_{t_i} = \mathbf{VS}(t_i) - \mathbf{H}(\hat{X}_{t_i|t_{i-1}})$ is the five-dimensional vector of time-t_i variance swap rate prediction errors, which in view of (4.1) is distributed as $D_x\mathbf{H}(\hat{X}_{t_i|t_{i-1}})\,(X_{t_i} - \hat{X}_{t_i|t_{i-1}}) + \eta_{t_i}$ with covariance matrix $V_{t_i|t_{i-1}}$, and $\mathcal{N} = 2{,}832$ is the sample size of daily observations.

It is known that univariate affine models cannot capture the empirical features of variance swap rates, e.g., Egloff, Leippold, and Wu (2010), and Aït-Sahalia, Karaman, and Mancini (2012). These models, for example, can only produce upward or downward sloping term structures, and variance swap rates have all the same persistence. Such model-based features of variance swap rates are in sharp contrast with the empirical features summarized in Table 12. In principle our univariate quadratic model in Section 3.1 could capture these features. Intuitively, the quadratic structure of the spot variance v_t relaxes the constraints imposed by an affine specification and is to some extent similar to a bivariate affine structure, when the two factors (X_t and X_t^2) are tightly related to each other.

We begin model estimations by fitting each of the three canonical representations of the univariate quadratic model in Section 3.1 to the variance swap data. We find that the largest log-likelihood of the univariate quadratic model is achieved when the state process X_t is in Class 3 (Proposition 3.2). This finding is confirmed by Akaike and Bayesian Information Criteria (AIC and BIC).[36] Table 13 reports the corresponding parameter estimates. The model parameters are estimated rather imprecisely, as their robust standard errors are fairly large. This may suggest that the univariate quadratic model is overparameterized, in the sense that it has too many parameters to fit available variance swap data and cannot be estimated precisely. In that case, imposing certain parameter restrictions should not deteriorate the fitting significantly. We consider four parametric restrictions that induce four alternative model specifications. Each restriction is tested via a likelihood ratio (LR) test.[37]

Specification 1 imposes that X_t has an affine dynamic by setting the quadratic coefficient $A = 0$ in (3.7). Specification 2 constrains the spot variance function, $v_t = \phi + \psi X_t + \pi X_t^2$, to be linear in X_t

[36] When the state process X_t is in Class 1, 2, and 3, AIC are $-97{,}316$, $-97{,}310$ and $-97{,}346$, and BIC are $-97{,}268$, $-97{,}262$ and $-97{,}298$, respectively. Both criteria achieved the minimum value when X_t is in Class 3.

[37] Denote L_U the likelihood of the unrestricted model and L_R the likelihood of the restricted model. Under the null hypothesis that the restriction holds true in the data generating process, the likelihood ratio statistic, $2\log(L_U/L_R)$, has asymptotically a chi-square distribution with degrees of freedom equal to the number of restrictions. If the null hypothesis were marginally rejected, the outcome of the test would have to be interpreted cautiously, as (4.2) is the quasi log-likelihood. However, as discussed below, the four parametric restrictions are strongly rejected by LR tests.

by setting $\pi = 0$. The corresponding LR tests strongly reject both restrictions, suggesting that the quadratic features of X_t and v_t play an important role in fitting variance swap rates.

Specification 3 restricts the functional form of the spot variance by imposing the spot variance function to have exactly one root, i.e., $\psi^2 = 4\phi\pi$. This guarantees the nonnegativity of the spot variance for any realization of X_t. Specification 4 further restricts Specification 3 by testing whether the root is at $X_t = 0$, i.e., $\phi = \psi = 0$. The corresponding LR tests strongly reject both restrictions, confirming that a flexible quadratic link between v_t and X_t is statistically important to fit variance swap rates.[38] To summarize, these statistical tests suggest that the full flexibility of the univariate quadratic model is necessary to fit variance swap rates.

We now investigate whether enriching the functional form of the spot variance can improve the fitting of the data. We estimate the univariate polynomial variance swap model in Section 3.3 when the state process X_t follows a quadratic diffusion and the degree of the polynomial is $N = 5$. The choice $N = 5$ asserts that the univariate polynomial model has the same number of parameters as the bivariate quadratic model, estimated next. Table 13 reports the parameter estimates.[39] The additional parameters, p_3, p_4, p_5, allow for a modest increase in the log-likelihood, which is not statistically significant according to a LR test, and a modest reduction of the AIC and BIC. Moreover, model parameters are still estimated quite imprecisely, according to robust standard errors. Thus, the polynomial form of the spot variance helps only marginally to improve the fitting of variance swap rates.

We now turn to the bivariate extension of the quadratic model in Section 3.2. We estimate all the identifiable equivalence class combinations of X_{1t} and X_{2t}, and find that the best fit, in terms of likelihood, AIC and BIC, is obtained when X_{1t} is in Class 1 and X_{2t} is in Class 3. Table 13 reports the parameter estimates, as well as AIC and BIC values. Interestingly, nearly all parameters are estimated very precisely, as can be seen from the small robust standard errors.

The log-likelihood of the bivariate model is significantly larger than the log-likelihoods of univariate models and the values of the BIC and AIC are significantly lower. The LR statistic of the bivariate model versus the univariate quadratic model is 27,980. The Vuong (1989) statistic of the bivariate model versus the univariate polynomial model is 46.3. These statistics are both highly significant and strongly reject the null hypothesis that the bivariate quadratic model is equivalent to any of the other two models.[40]

Following Giacomini and White (2006), we also compare the bivariate model and the univariate models using scoring-type rules. The test statistic is the log-likelihood under the bivariate model

[38] Under the four null hypotheses, namely, 1) $A = 0$, 2) $\pi = 0$, 3) $\psi^2 = 4\phi\pi$ and 4) $\phi = \psi = 0$, the LR test statistics are 386, 620, 264, 286, respectively, and are all well above any conventional critical value.

[39] The relation between model parameters in Section 3.3 and those in Table 13 is straightforward, namely $p_0 = \phi$, $p_1 = \psi$ and $p_2 = \pi$.

[40] The asymptotic distribution of the test statistics under the null hypotheses are the chi-square with 5 degrees of freedom and standard normal, respectively. Recall that the bivariate quadratic model nests the univariate quadratic model. Setting $b_2 = \beta_{22} = a_2 = \alpha_2 = A_2 = 0$ in the bivariate model, i.e., imposing 5 parameter restrictions, implies that X_{2t} is constant and can be normalized to 1 for identification purposes. Thus, β_{12} in the bivariate model parametrization corresponds to b_1 in the univariate model parametrization.

minus the log-likelihood under the univariate quadratic or polynomial model. If the two models are equivalent, the test statistic has zero mean, which can be tested using a simple t-test.[41] The t-statistics are 10.6 and 9.8, respectively, and are both highly significant. This further supports that the bivariate quadratic model fits variance swap rates significantly better than the univariate models.

Finally, Figure 17 shows the filtered trajectories of the state process X_t in the bivariate model. It suggests a natural interpretation of its components. X_{1t} is more volatile and mimics the time series trajectories of the short term variance swap rates, mainly capturing sudden movements in those rates. X_{2t} is more persistent and mainly captures long term movements in variance swap rates.

4.3 Goodness-of-fit Tests

To corroborate the above likelihood-based analysis, we now discuss the variance swap pricing errors for the three models and run various goodness-of-fit tests.

Table 14 summarizes the pricing errors, which are defined as model-based minus actual variance swap rates, both in volatility units. Consistently with the likelihood-based analysis, the bivariate quadratic model nearly always outperforms the other models in terms of bias and root mean square error (RMSE), and often to a large extent. For example, in the out-of-sample period, the RMSE of the bivariate quadratic model for the 2-month variance swap rates is 65% lower than the RMSE of the other models. The comparison between the bivariate quadratic model and the univariate polynomial model is particularly interesting, as the two models have the same number of parameters. In most cases, the RMSE of the bivariate model is less than half the RMSE of the polynomial model, both in-sample and out-of-sample.

Figure 18 shows actual and model-based trajectories under the bivariate quadratic model of the 2-month and 2-year variance swap rates, which are respectively the most and least volatile rates. The good performance of the model is evident throughout the in-sample and out-of-sample periods. A small lack of fit of the highest values of the 2-year variance swap rates is noticeable in the out-of-sample period, which includes the market turmoils of Fall 2008.

To assess the statistical differences of pricing errors of the different models, we run various Diebold–Mariano (DM) tests.[42] For each model and each term, the time-t loss function is given by the absolute pricing error, $L(e_t) = |e_t|$, where $e_t = \sqrt{G(\tau, X_t)/\tau} - \sqrt{\mathrm{VS}(t, t+\tau)}$.[43] Denote the time-$t$ loss differential between the univariate and bivariate quadratic models by $d_t^{(u,b)} = L(e_t^{(u)}) - L(e_t^{(b)})$. The loss differential between the polynomial and bivariate models, $d_t^{(p,b)}$, is similarly defined. Under the null hypothesis that the two models have pricing errors of equal magnitude, $\mathbb{E}_{\mathbb{P}}[d_t^{(u,b)}] = 0$. If the bivariate model outperforms the univariate model, then $\mathbb{E}_{\mathbb{P}}[d_t^{(u,b)}] > 0$. The DM statistic is the t-

[41] We view this test as a main robustness check of the previous LR and Vuong's tests. Given the autocorrelation and heteroscedasticity in the log-likelihood differences, robust standard errors are computed using the Newey and West (1987) variance estimator with the number of lags optimally chosen according to Andrews (1991).

[42] We follow the standard practice in the literature of using Diebold–Mariano tests to draw conclusions about models, rather than about model forecasts; see Diebold (2012) for a discussion of this point.

[43] Note that the time-t pricing error considered here uses the time-t filtered value of X_t, not its prediction as in (4.2), which makes the DM tests complementary to the likelihood-based analysis in the previous section.

statistic for this test.[44] Table 14 reports the results. DM tests strongly confirm that the bivariate model significantly outperforms the univariate quadratic and polynomial models.[45] As a robustness check, we also run DM tests using pricing errors in variance units, rather than volatility units, i.e., $e_t = G(\tau, X_t)/\tau - \mathrm{VS}(t, t+\tau)$, and using quadratic loss functions, rather than absolute loss functions. These additional DM tests strongly confirm the results in Table 14.

Finally, we run predictive regressions for each model and each term. We regress the actual future variance swap rate $\mathrm{VS}(t, t+\tau)$ on a constant and the d-day ahead, model-based prediction, $\mathbb{E}_{\mathbb{P}}[G(\tau, X_t)/\tau | \mathcal{F}_{t-d}]$, obtained at time $t-d$, i.e.,

$$\mathrm{VS}(t, t+\tau) = \gamma_0 + \gamma_1 \, \mathbb{E}_{\mathbb{P}}[G(\tau, X_t)/\tau | \mathcal{F}_{t-d}] + \mathrm{error}_t.$$

If the model captures well the variance swap term structure dynamics, then it should provide unbiased, $\gamma_0 = 0$, and efficient, $\gamma_1 = 1$, forecasts of future variance swap rates. As a benchmark, we consider the martingale model that uses the actual variance swap rate at time $t-d$ as a predictor of the future variance swap rate. This model is a challenging benchmark because of the strong persistence of variance swap rates; first order autocorrelations of variance swap rates range from 0.984 to 0.995, Section 4.1. We consider two forecasting horizons, $d = 1$ day and $d = 10$ days. Table 15 reports the regression results.[46] Interestingly, for both forecasting horizons and nearly all terms, the bivariate quadratic model provides unbiased and efficient variance swap rate forecasts, as can be seen from the high p-values of the null hypotheses $H_0 : \gamma_0 = 0$ and $H_0 : \gamma_1 = 1$. Only the bivariate quadratic model passes all these tests and outperforms the martingale model. The univariate quadratic and polynomial models provide biased and inefficient forecasts in most cases. The martingale model provides relatively accurate forecasts for the 1-day horizon, but its forecasting accuracy deteriorates when moving to the 10-day horizon. The univariate quadratic model provides the least accurate forecasts. To summarize, also predictive regressions strongly confirm that the bivariate quadratic model captures well the variance swap term structure dynamics, even outperforming the martingale model.

5 Optimal Portfolios: Theoretical Setup

In this section, we formalize and solve an optimal portfolio problem for variance swaps, stock index, and risk free bond. As at the beginning of Section 3, we consider a diffusion process X_t in some state space $\mathcal{X} \subset \mathbb{R}^m$, solving the SDE (3.1) where W_t is a standard d-dimensional Brownian motion under the risk neutral measure $\mathbb{Q}$. The spot variance, v_t, and variance swap rates, $\mathrm{VS}(t, T)$, are given as functions of the state variable, X_t, by (3.4) and (3.5), respectively.

[44] The standard error is computed using the Newey and West (1987) autocorrelation and heteroscedasticity consistent variance estimator with the number of lags optimally chosen according to Andrews (1991).

[45] The DM test statistics are positive but not significant only for the 6 month and 1 year variance swaps in the out-of-sample period, which may be due to the limited sample size, i.e., 794 daily observations.

[46] Also in these regressions, robust standard errors are computed using the Newey and West (1987) covariance matrix estimator with the number of lags optimally chosen according to Andrews (1991). Given the strong persistence of variance swap rates, all R^2 of predictive regressions are high, between 70% and 99%, and not reported.

5.1 Investing in Variance Swaps

We compute the return of an investment in variance swaps. Fix a term $\tau > 0$, and consider a τ-variance swap issued at some inception date t^*. Denote its maturity $T^* = t^* + \tau$. The nominal spot value Γ_t at date $t \in [t^*, T^*]$ of a one dollar notational long position in this variance swap is given by

$$\begin{aligned}\Gamma_t &= \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-r(T^*-t)}\frac{1}{\tau}\left(\int_{t^*}^{T^*} v_s\,ds - \tau \mathrm{VS}(t^*, T^*)\right) \mid \mathcal{F}_t\right] \\ &= \frac{\mathrm{e}^{-r(T^*-t)}}{\tau}\left(\int_{t^*}^{t} v_s\,ds + (T^*-t)\mathrm{VS}(t, T^*) - \tau \mathrm{VS}(t^*, T^*)\right)\end{aligned}$$

where r is the constant risk free rate. In stochastic differential form, we obtain $d\Gamma_t = \Gamma_t\, r\, dt + dM_t$ with the $\mathbb{Q}$-martingale increment excess return

$$dM_t = \frac{\mathrm{e}^{-r(T^*-t)}}{\tau}\left(v_t\,dt + d\left((T^*-t)\mathrm{VS}(t, T^*)\right)\right) = \frac{\mathrm{e}^{-r(T^*-t)}}{\tau}\nabla_x G(T^*-t, X_t)^\top \Sigma(X_t)\,dW_t$$

where ∇_x denotes the gradient. Now fix a date $t \in [t^*, T^*)$, and consider an investor with positive wealth V_t who takes a position in this variance swap with relative notional exposure of n_t. The cost of entering such a position is $n_t V_t \Gamma_t$. The remainder of the wealth, $V_t - n_t V_t \Gamma_t$, is invested in the risk free bond. This makes the investment self-financing. At a later instant $t + dt$, the wealth has grown to $V_{t+dt} = (V_t - n_t V_t \Gamma_t)(1 + r\,dt) + n_t V_t \Gamma_{t+dt}$. The resulting rate of return is

$$\frac{dV_t}{V_t} = \frac{V_{t+dt} - V_t}{V_t} = (1 - n_t\Gamma_t)\,r\,dt + n_t\,d\Gamma_t = r\,dt + n_t\,dM_t.$$

Consider now τ-variance swaps that are issued at a sequence of inception dates $0 = t_0^* < t_1^* < \cdots$, with $t_{k+1}^* - t_k^* \le \tau$, for example 3-month variance swaps issued every month. At any date $t \in [t_k^*, t_{k+1}^*)$ the investor takes a position in the respective on-the-run τ-variance swap with maturity $T^*(t) = t_k^* + \tau$. In the limit case where a new τ-variance swap is issued at any date t, we obtain a "sliding" variance swap investment, and we set $T^*(t) = t + \tau$. Iterating the above reasoning shows that the resulting wealth process V_t evolves according to

$$\frac{dV_t}{V_t} = r\,dt + n_t\,\frac{\mathrm{e}^{-r(T^*(t)-t)}}{\tau}\nabla_x G(T^*(t)-t, X_t)^\top \Sigma(X_t)\,dW_t \tag{5.1}$$

where the excess return on the right hand side is a $\mathbb{Q}$-martingale increment.

5.2 Optimal Portfolio Problem

We consider an investment universe consisting of the risk free bond, stock index S, and n on-the-run variance swaps with different terms $\tau_1 < \cdots < \tau_n$ and respective issuance dates encoded by n maturity

functions $T_1^*(t), \ldots, T_n^*(t)$, as defined above. The stock index price process has $\mathbb{Q}$-dynamics

$$\frac{dS_t}{S_t} = r\,dt + \sqrt{g(X_t)}\,\mathbf{R}(X_t)^\top dW_t \tag{5.2}$$

where $\mathbf{R} = (R_1, \ldots, R_d)^\top : \mathcal{X} \to \mathbb{R}^d$ is some function with constant norm $\|\mathbf{R}\| \equiv 1$, modeling the correlation between stock returns and spot variance changes.[47]

Let w_t denote the fraction of wealth invested in the stock index and $\mathbf{n}_t = (n_{1t}, \ldots, n_{nt})^\top$ the vector of relative notional exposures to each on-the-run τ_i-variance swap, $i = 1, \ldots, n$. To make the investment self-financing, the fraction of wealth invested in the risk free bond is given by $1 - w_t - \mathbf{n}_t^\top \mathbf{\Gamma}_t$, where $\mathbf{\Gamma}_t$ is the vector of the variance swap spot values. Combining (5.1) and (5.2), the resulting wealth process V_t has $\mathbb{Q}$-dynamics

$$\frac{dV_t}{V_t} = r\,dt + \left(\mathbf{n}_t^\top, w_t\right) \mathcal{G}(t, X_t)\,dW_t \tag{5.3}$$

with the $(n+1) \times d$-volatility matrix $\mathcal{G}(t, X_t)$ defined by

$$\mathcal{G}(t,x) = \begin{pmatrix} \mathcal{D}(t,x) & \mathbf{0}_{n\times 1} \\ \mathbf{0}_{1\times m} & \sqrt{g(x)} \end{pmatrix} \begin{pmatrix} \Sigma(x) \\ \mathbf{R}(x)^\top \end{pmatrix} \tag{5.4}$$

where $\mathcal{D}(t,x)$ is the $n \times m$-matrix whose ith row is given by $\left(\mathrm{e}^{-r(T_i^*(t)-t)}/\tau_i\right) \nabla_x G(T_i^*(t) - t, x)^\top$.

We now formulate the optimal portfolio problem. We fix a finite time horizon T, and maximize expected utility from terminal wealth of an investor with power utility function $u(V) = V^{1-\eta}/(1-\eta)$ and constant relative risk aversion $\eta > 0$. That is, we solve the optimization problem

$$\max_{\mathbf{n}, w} \mathbb{E}_{\mathbb{P}}\left[u(V_T)\right] \tag{5.5}$$

for some given initial wealth V_0. The investor takes the market price of risk as given, which we specify as follows.[48] The objective probability measure $\mathbb{P}$ is related to the risk neutral measure $\mathbb{Q}$ on $\mathcal{F}_T$ via the Radon–Nikodym density

$$\frac{d\mathbb{Q}}{d\mathbb{P}}\,|_{\mathcal{F}_T} = \exp\left(-\int_0^T \Lambda(X_t)^\top dW_t + \frac{1}{2}\int_0^T \|\Lambda(X_t)\|^2\,dt\right) \tag{5.6}$$

for some sufficiently regular market price of risk function $\Lambda : \mathcal{X} \to \mathbb{R}^d$.[49]

[47] The price process dynamic in (5.2) is tantamount to $dS_t/S_t = r\,dt + \sqrt{v_t}\,dB_t$ for the scalar $\mathbb{Q}$-Brownian motion B_t defined as $dB_t = \mathbf{R}(X_t)^\top dW_t$. That is, B_t and W_t have correlation $d\langle B, W_k\rangle_t/dt = R_k(X_t)$.

[48] By exogenously specifying the market price of risk we take a partial equilibrium view. This approach is standard in the optimal allocation literature, e.g., Liu and Pan (2003), Chacko and Viceira (2005), and Liu (2007).

[49] If we denote the corresponding Girsanov transformed $\mathbb{P}$-Brownian motion by $dW_t^{\mathbb{P}} = dW_t - \Lambda(X_t)\,dt$, we obtain the familiar stochastic exponential representation

$$\frac{d\mathbb{Q}}{d\mathbb{P}}\,|_{\mathcal{F}_T} = \exp\left(-\int_0^T \Lambda(X_t)^\top dW_t^{\mathbb{P}} - \frac{1}{2}\int_0^T \|\Lambda(X_t)\|^2\,dt\right).$$

Since the number n of on-the-run variance swaps available in the market can be chosen arbitrarily large, it is no essential loss in generality to assume market completeness.

Assumption 5.1. *The market is complete with respect to the stock index and the n on-the-run τ_i-variance swaps. Specifically, we assume that the filtration $\mathcal{F}_t = \mathcal{F}_t^W$ is generated by W_t, and that the $(n+1) \times d$-volatility matrix $\mathcal{G}(t, X_t)$ is injective $dt \otimes d\mathbb{Q}$-a.s.*

Appendix 8 shows that, as a consequence of Assumption 5.1, the dimension d of the Brownian motion cannot exceed the number m of factors and the number n of variance swaps by more than one, i.e., $d \leq m+1$ and $d \leq n+1$. Moreover, the maturity date functions $T_i^*(t)$ have to be mutually different for all t.

We now state the existence and characterization result for the optimal strategy with standard technical assumptions and proof given in Appendix 7.

Theorem 5.1. *Under Assumptions 5.1 and 7.1–7.3 in Appendix 7 there exists an optimal strategy $\mathbf{n}_t^*, w_t^*$ given as solution of the linear equation*

$$\mathcal{G}(t, X_t)^\top \begin{pmatrix} \mathbf{n}_t \\ w_t \end{pmatrix} = \frac{1}{\eta} \Lambda(X_t) + \Sigma(X_t)^\top \nabla_x h(T-t, X_t) \tag{5.7}$$

where the function $h(\tau, x)$ is defined in (7.3) *in Appendix 7.*

The optimal strategy is thus composed of the familiar myopic and intertemporal hedging terms, as discussed in Merton (1971). The myopic demand, coming from $\Lambda(X_t)/\eta$, would be the mean-variance optimal investment over the next instant not accounting for future investments, or assuming a constant investment opportunity set. The intertemporal hedging demand, coming from $\Sigma(X_T)^\top \nabla_x h(T-t, X_t)$, arises due to the need to hedge against fluctuations in the investment opportunities. These fluctuations are induced, inter alia, by the stochastic volatility of the stock index. We discuss the computation of $\nabla_x h(T-t, X_t)$ in Appendix 9.

The following corollary shows that variance swaps can be used to span volatility risk. The optimal investment in the stock index is thus only seeking its risk premium. In their affine setting, Egloff, Leippold, and Wu (2010) reach the same conclusion. Corollary 5.1 extends this result to a general multivariate diffusion setting. The proof is given in Appendix 8.

Corollary 5.1. *If $d > m$ then the optimal investment in the stock index, w_t^*, is fully determined by the myopic term and does not depend on the choice of the variance swaps.*

5.3 Bivariate Quadratic Model Specification

We now resume the bivariate quadratic variance model in Section 3.2. Our empirical analysis in Section 4 shows that the best fit is attained when X_{1t} is in Class 1 and X_{2t} is in Class 3. We focus on this specification in the following. The dimension of the Brownian motion W_t is $d = 3$, and the

2×3-dispersion matrix $\Sigma(x)$ takes the form

$$\Sigma(x) = \begin{pmatrix} \sqrt{1 + A_1 x_1^2} & 0 & 0 \\ 0 & \sqrt{x_2 + A_2 x_2^2} & 0 \end{pmatrix}.$$

To account for the widely documented correlation between index returns and spot variance changes, e.g., Broadie, Chernov, and Johannes (2007), and Aït-Sahalia and Kimmel (2007), the correlation vector function is chosen to be of the form $\mathbf{R}(x) = \left(R_1(x), 0, \sqrt{1 - R_1(x)^2}\right)^\top$. The correlation between index returns and variance changes is then given by

$$\text{Corr}\left(\frac{dS_t}{S_t}, dv_t\right) = \frac{\nabla_x g(X_t)^\top \Sigma(X_t)}{\|\nabla_x g(X_t)^\top \Sigma(X_t)\|} \mathbf{R}(X_t) = \text{sign}\left(\psi + 2\pi X_{1t}\right) R_1(X_t).$$

We set $R_1(x) = -\text{sign}\left(\psi + 2\pi x_1\right) \times 0.7$ to achieve a constant correlation of -0.7, in line with the literature. As a consequence, we obtain $R_3(x) = \sqrt{1 - 0.7^2} = 0.714$.

Consistently with (3.8), we specify the market price of risk function as

$$\Lambda(x) = \left(\frac{\lambda_0 + \lambda_1 x_1}{\sqrt{1 + A_1 x_1^2}}, 0, \Lambda_3(x)\right)^\top \tag{5.8}$$

where $\Lambda_3(x)$ is implicitly defined, up to its sign, by

$$\Lambda_3(x) = \pm\sqrt{\|\Lambda(x)\|^2 - \Lambda_1(x)^2}. \tag{5.9}$$

The sign of $R_3(x)\Lambda_3(x)$ has a direct impact on the equity risk premium, which is given by

$$\frac{\mathbb{E}_\mathbb{P}[dS_t/S_t \mid \mathcal{F}_t] - \mathbb{E}_\mathbb{Q}[dS_t/S_t \mid \mathcal{F}_t]}{dt} = \sqrt{g(X_t)}\mathbf{R}(X_t)^\top \Lambda(X_t). \tag{5.10}$$

Based on our estimations, $R_3(X_t)\Lambda_3(X_t)$ is much larger in absolute value than $R_1(X_t)\Lambda_1(X_t)$. Since $R_3(x)$ is positive, a negative $\Lambda_3(x)$ would lead to a negative equity risk premium, which would be economically odd, so we take the positive square root in (5.9). Clearly, $\|\Lambda(x)\|^2$ needs to be specified so that the argument in the square root in (5.9) is nonnegative for all $x \in \mathcal{X}$. We specify it as proportional to spot variance

$$\|\Lambda(x)\|^2 = \kappa\, g(x) \tag{5.11}$$

with $\kappa \geq \kappa^* = \max_{x \in \mathcal{X}} \Lambda_1(x)^2 / g(x)$.[50] Since $\Lambda_1(x)$ is uniformly bounded in x, it follows that the spot variance $g(x)$ and the equity risk premium (5.10) are increasing functions in x_1, for x_1 large enough. This means that the equity risk premium increases in bad times, i.e., when variance increases and stock index falls due to the leverage effect. Such a countercyclical equity risk premium is certainly a desirable feature of our model and motivates the chosen specification (5.11) of $\|\Lambda(x)\|^2$.

[50] Alternatively, we could specify $\|\Lambda(x)\|^2 = c$, for some constant $c \geq \max_{x \in \mathcal{X}} \Lambda_1(x)^2$. This specification implies that $\nabla_x h(\tau, x) = 0$, because the function h defined in (7.3) in Appendix 7 no longer depends on x. Hence, in this case the optimal investment in variance swaps and stock index in (5.7) consists of myopic demand alone and is available in closed form.

We set $\kappa = 1.58$ in (5.11), which corresponds to a sample average equity risk premium of 6%.[51] Figure 19 shows the induced model-based time series of equity risk premium (5.10), which exhibits significant time variation and countercyclical behavior. Figure 19 also shows the induced time series of variance risk premium, which is given by

$$\frac{\mathbb{E}_{\mathbb{P}}[dv_t \mid \mathcal{F}_t] - \mathbb{E}_{\mathbb{Q}}[dv_t \mid \mathcal{F}_t]}{dt} = (\psi + 2\pi X_{1t})(\lambda_0 + \lambda_1 X_{1t}). \tag{5.12}$$

The variance risk premium is procyclical, takes both positive and negative values, and is negative most of the time. Both model-based equity and variance risk premiums are economically sizable and follow plausible dynamics which are in line with recent studies, e.g., Bollerslev and Todorov (2011), Aït-Sahalia, Karaman, and Mancini (2012), and Martin (2013). These features lend further empirical support to our quadratic variance swap model.

5.4 Optimal Portfolios in the Bivariate Quadratic Model

We assume that $n = 2$ variance swaps are available for investment, specified by their maturity date functions $T_1^*(t)$ and $T_2^*(t)$. We allow for various roll-over strategies. In all cases the maturity date functions differ, $T_1^*(t) \neq T_2^*(t)$, for all t, which is important in view of Assumption 5.1. It is a tedious but routine exercise to check that all assumptions underpinning Theorem 5.1 and Corollary 5.1 are satisfied. Appendix 10 sketches the arguments.

The optimal fraction of wealth invested in the stock index is given by

$$w_t^* = \frac{\Lambda_3(X_t)}{\eta\sqrt{g(X_t)}R_3(X_t)} \tag{5.13}$$

which is recovered by setting $\mathbf{v} = (0, 0, 1)^\top$ in (8.2) in Appendix 8. As stated in Corollary 5.1, it is fully determined by the myopic term and does not depend on the choice of the variance swaps. It follows that the optimal weight w_t^*, while being state-dependent, is uniformly bounded from below and above with sharp bounds given by

$$\frac{\sqrt{\kappa - \kappa^*}}{\eta\sqrt{1 - 0.7^2}} \leq w_t^* \leq \frac{\sqrt{\kappa}}{\eta\sqrt{1 - 0.7^2}}. \tag{5.14}$$

The intertemporal hedging demand is fully borne by the optimal investment in the variance swaps. Plugging (5.13) in (5.7) shows that the optimal vector of relative notional exposures to the respective on-the-run variance swaps is given as solution $\mathbf{n}_t^* = \mathbf{n}_t$ of the linear equation

$$\Sigma(X_t)^\top \mathcal{D}(t, X_t)^\top \mathbf{n}_t = \frac{1}{\eta}\left(\Lambda(X_t) - \frac{\Lambda_3(X_t)}{R_3(X_t)}\mathbf{R}(X_t)\right) + \Sigma(X_t)^\top \nabla_x h(T - t, X_t). \tag{5.15}$$

[51] The equity risk premium is notoriously difficult to estimate. Merton (1980) even argues that a positive risk premium should be explicitly modeled, and various studies have followed this approach, e.g., Jackwerth (2000), and Barone-Adesi, Engle, and Mancini (2008).

We provide a closed form approximation of $\nabla_x h(T-t, X_t)$ in (10.1) in Appendix 10.

6 Optimal Portfolios: Empirical Findings

We perform an empirical analysis of optimal portfolios in the above bivariate quadratic model. The investment universe consists of the risk free bond, the stock index, and 3-month and 2-year variance swaps, rolled over monthly and yearly, respectively. The initial wealth is normalized to 100. The investment horizon is $T = 14.4$ years, which is the time span of our sample. The risk aversion is set to $\eta = 5$, which is an average value in survey data.[52] For a comparison we also consider $\eta = 1$, which corresponds to logarithmic utility. When the log-investor has only access to the stock index and the risk free bond, it is well known that the optimal weight in the stock index is given by the ratio of equity risk premium and spot variance, $\mathbf{R}(X_t)^\top \Lambda(X_t)/\sqrt{g(X_t)}$, e.g., Filipović and Platen (2009). Optimal portfolios are rebalanced daily. That is, each day optimal portfolio weights are adjusted according to (5.13) and (5.15). We also consider proxy portfolios with lower rebalancing frequencies. We first study the optimal portfolios using historical data. We then perform a Monte Carlo analysis of optimal portfolios. Section 6.3 discusses several robustness checks that largely confirm our results.

6.1 Optimal and Proxy Portfolios

Figures 20 and 21 display the optimal portfolio weights in the stock index, and in on-the-run 3-month and 2-year variance swaps, for $\eta = 5$ and $\eta = 1$, respectively. The optimal weights in variance swaps follow a short-long strategy, with a short position in the 2-year variance swap and a long position in the 3-month variance swap. As the negative variance risk premium in 2-year variance swaps is larger in absolute value than the risk premium in 3-month variance swaps (Section 4.1), going short in 2-year variance swaps allows to reap the large risk premium. Short positions in 2-year variance swaps are partially hedged via long positions in 3-month variance swaps, limiting portfolio losses when volatility increases. The 3-month variance swap is more sensitive to volatility increases than the 2-year variance swap, and is thus an effective hedging instrument.

The optimal weights in variance swaps exhibit significant periodic patterns, with increasing weights in absolute value when their maturities are approaching. Intuitively, close to maturity, most realized variance has accumulated, inducing little volatility in spot value and thus reducing the risk premium carried by the variance swap. To keep an optimal level of portfolio risk exposure and earn risk premiums, the optimal weights in variance swaps need to increase in absolute value.

The optimal weight in the stock index (5.13) is positive, which is consistent with the positive equity risk premium to be earned. In contrast to the weights in variance swaps, the stock index weight exhibits no periodic pattern, which is in line with Corollary 5.1. The bounds (5.14) have a stabilizing effect on the optimal fraction of wealth invested in the stock index. The optimal weights in the stock

[52] Meyer and Meyer (2005) survey some of the key studies by economists of how the coefficient of relative risk aversion varies across the population. Most of the survey data suggests values between 0.23 and 8.

index and the 3-month variance swap are significantly larger for $\eta = 1$ than for $\eta = 5$. The log-investor seeks significantly more exposure to the stock index and to a lesser extent to the 2-year variance swap, and needs a larger position in the 3-month variance swap to hedge this exposure.

Some oscillations in portfolio weights are observed during the low volatility period 2005–2006. Because volatility reaches historically low values, variance swap rates are also low. This renders the matrix $\mathcal{D}(t, X_t)$ in equation (5.15) for $\mathbf{n}_t$ close to singular. However, low volatility also implies small actual returns in the stock index and variance swaps. This in turn annihilates the impact of oscillating portfolio weights on the wealth process, resulting in non-oscillating wealth trajectories, as shown below in Figures 23 and 24.

In view of Corollary 5.1, the optimal weight in the stock index includes myopic demand only. Figure 22 decomposes the optimal weights in variance swaps into myopic and intertemporal hedging demands according to (5.7), for $\eta = 5$. Myopic and intertemporal hedging demands are positive for the 3-month, and negative for the 2-year variance swaps, reflecting the risk premium versus hedging trade-off in these contracts. The periodic patterns in the optimal weights in variance swaps are mainly borne by the intertemporal hedging demand, while the myopic demand only exhibits little periodicity. The hedging demand gets closer to zero as the terminal investment date approaches. As the investment horizon shrinks, the need to hedge against future fluctuations in the investment opportunity set becomes less important and the investor behaves more myopically. For the log-optimal portfolio, $\eta = 1$, it is well known that the intertemporal hedging demand is zero.[53]

Figure 23 shows the wealth trajectory of the optimal portfolio for $\eta = 5$. The wealth trajectory exhibits low volatility and steady growth. This results in a Sharpe ratio of 1.46%, which is larger than the Sharpe ratio of 1.20% of the S&P 500. Thus, optimally investing in variance swaps and stock index allows for a smooth wealth growth, which is far less sensitive to market falls than investing in the stock index only. The S&P 500 yields a higher terminal wealth than the optimal portfolio. Indeed, the optimal portfolio is not designed to maximize terminal wealth. Compared to the stock index, the optimal portfolio may exhibit lower returns on some occasions but it has always a lower volatility. This implies that including variance swaps in the portfolio of a risk averse investor brings more utility than investing in the stock index only.

Figure 24 shows the wealth trajectories of the optimal portfolios with and without variance swaps for $\eta = 1$. The log-optimal wealth process including variance swaps has a Sharpe ratio of 1.54%, and exhibits significantly larger fluctuations than the S&P 500, which contrasts with the optimal wealth trajectory of the more risk averse investor with $\eta = 5$. This suggests that variance swaps can be used either to seek additional risk premiums or achieve stable wealth growth, depending on the risk profile of the investor. The trajectory of the log-optimal portfolio in stock and bond is very similar to the S&P 500, as the optimal weight in the stock index turns out to be close to one.

We now study the performance of proxy portfolios when the number of contracts in the portfolio is rebalanced at lower frequencies than daily. Specifically, the stock index and 3-month variance swap

[53] It follows from (7.3) in Appendix 7 that the function $h(\tau, x)$ in equation (5.7) is zero for $\eta = 1$.

positions are rebalanced monthly, and the 2-year variance swap position is rebalanced yearly. Between rebalancing dates, positions are kept constant. At rebalancing dates t^*_{ik}, $i = 1, 2$, variance swap investments are rolled over to newly issued 3-month and 2-year variance swaps, respectively, according to the portfolio weights $\overline{n}_{it^*_{ik}}$ given as exponentially weighted average of past optimal portfolio weights,

$$\overline{n}_{it^*_{ik}} = \frac{\sum_{t^*_{i,k-1}<t\leq t^*_{ik}} n^*_{it}\,\omega_{it}}{\sum_{t^*_{i,k-1}<t\leq t^*_{ik}} \omega_{it}}$$

where $\omega_{it} = \mathrm{e}^{-(t^*_{ik}-t)}$.[54] These portfolio weights attempt to capture the periodic pattern of the optimal weights over the lifetime of the variance swaps. The rationale for assessing the performance of this proxy portfolio is twofold. First, low rebalancing frequencies obviously reduce transaction costs when implementing the portfolio strategy in practice. Second, the portfolio gains can be evaluated using market data, without resolving to model-based variance swap rates.[55] Figures 23 and 24 show that the wealth trajectories of these proxy portfolios are similar to the ones of the optimal portfolios, for $\eta = 5$ and $\eta = 1$. This is rather remarkable. Although this is mainly an in-sample result based on one historical realization, it suggests that our optimal portfolio strategies have potential to be implemented in practice.

The results documented above differ from those in Egloff, Leippold, and Wu (2010) in a number of ways. In their affine setting, the optimal weight in the stock index is constant over time and the optimal weights in variance swaps are state-independent. In our quadratic setting, optimal portfolio weights depend on state variables and exhibit the rich dynamics discussed above. Thus, the two optimal strategies are fundamentally different. Furthermore, they assume that at any time one can invest in newly issued variance swaps at zero spot value ("sliding" variance swap investment). This is a special case of our framework in which we take into account investments in on-the-run variance swaps. This allows us to uncover periodic patterns in the optimal variance swap weights. Moreover, their empirical implementation of optimal portfolios is static, while we implement dynamic strategies. They use a risk aversion of $\eta = 200$ while we use $\eta = 5$ and $\eta = 1$. Finally, market price of risk specifications are different in the two studies. This implies that optimal portfolio weights are significantly different and actually mirror each other.[56]

6.2 Monte Carlo Analysis

The optimal and proxy portfolios above are based on the historical realization of variance swap rates and the S&P 500. In this section we perform a Monte Carlo analysis of optimal and proxy portfolios. The goal is to evaluate their performance under different economic scenarios, and to assess the utility

[54] We set $\overline{n}_{i0} = n^*_{i0}$ for the initial holding period.

[55] The actual 2-month and 1-year variance swap rates needed to evaluate the gains in the 3-month and 2-year variance swaps at the rebalancing dates, respectively, are available in our dataset. The realized variance is given by the sum of squared daily log-returns of the S&P 500.

[56] As mentioned above, our optimal strategies take short positions in the long term variance swap (to earn the variance risk premium), long positions in the short term variance swap (to hedge volatility increases) and long positions in the stock index (to earn the equity risk premium). Egloff, Leippold, and Wu (2010) find opposite trading directions in their optimal strategy.

effect of a combined investment in variance swaps, stock index, and bond versus investing in stock index and bond, or holding the stock index only.

We simulate 10,000 trajectories of the bivariate quadratic model in Section 5.3, and measure the performance of the optimal and proxy portfolios. The investment horizon is $T = 2$ years. Table 16 reports certainty equivalent, implied rate, average Sharpe ratio, and average terminal wealth. The certainty equivalent is defined as the initial amount of money to be invested in the risk free bond that would yield the same terminal utility as the respective portfolio. The implied rate is defined as the constant annual rate of return on the initial wealth of 100 that would be needed to achieve the same terminal utility. Sharpe ratios are computed using daily changes of the portfolio value.

Optimal and proxy portfolios including variance swaps systematically outperform the stock index, as well as the log-optimal stock-bond portfolio. This holds true in terms of certainty equivalent and Sharpe ratio, irrespective of rebalancing frequency and risk aversion, highlighting the added economic value of variance swaps. The differences are economically important. For $\eta = 5$, the optimal portfolio has an implied rate of 2.53%. This is much higher than the respective implied rate of -2.73% of the stock index. For $\eta = 1$, the implied rate of the optimal portfolio including variance swaps is 4.34%, which is 61 and 64 basis points higher than the implied rate of the log-optimal stock-bond portfolio and of the stock index, respectively. Proxy portfolios have smaller certainty equivalents than the respective optimal portfolios, but still outperform the stock index, as well as the optimal stock-bond portfolio for $\eta = 1$.

Sharpe ratios of optimal and proxy portfolios including variance swaps are higher than Sharpe ratios of the stock index. This is interesting because the optimal portfolio is not designed to maximize the Sharpe ratio. It confirms our previous empirical findings, which were based on a single historical trajectory of variance swap rates and the stock index.

The expected terminal wealth corresponds to the utility of a risk neutral investor. Such an investor prefers the log-optimal portfolio including variance swaps, that yields an average terminal wealth of 114.06, over the stock index. In contrast, a risk averse investor with $\eta = 5$ prefers to include variance swaps in her portfolio at the cost of a lower expected terminal wealth of 106.08.

To summarize, variance swaps have a significant economic value for risk averse investors. Adding variance swaps to a portfolio improves its performance, also when rebalanced infrequently.

6.3 Robustness Checks

We performed several robustness checks that largely confirm our optimal portfolio results.

Optimal portfolios above are based on 3-month and 2-year variance swaps. Optimal portfolios based on variance swaps with other term combinations (such as 3-month and 1-year, 6-month and 1-year, 6-month and 2-year) have similar performance. The same holds when using different roll-over periods (such as daily, half term, or term of the variance swaps). For example, when the risk aversion is $\eta = 5$, the optimal wealth process always grows steadily over time and is significantly smoother than

the trajectory of the stock index. Indeed, since we are in a complete market setup, in theory, the choice of variance swap terms and roll-over periods has no impact on the optimal wealth trajectory. In particular, the optimal portfolio weight in the stock index neither depends on variance swap terms nor on roll-over periods, Corollary 5.1.

Besides the risk aversion levels of $\eta = 5$ and 1, we also experimented with higher values, such as $\eta = 30$. The optimal portfolio weights in the risky assets follow the same pattern. The weights are smaller in absolute value, which is consistent with the investor being more risk averse.

We also considered other investment horizons, such as 5 and 10 years. The pattern of optimal portfolio weights is only marginally affected by the choice of the investment horizon.

The above empirical analysis is based on a sample average equity risk premium of 6%. We redid the analysis for a sample average equity risk premium set to 4% by changing the parameter κ in (5.11) accordingly. This leads to smaller portfolio weights in the stock index, as theory predicts, and the pattern of the optimal weights in the variance swaps are essentially unaffected.

Risk aversion has a nonlinear impact on myopic and intertemporal hedging demands.[57] For the logarithmic utility case, $\eta = 1$, the intertemporal hedging demand is zero. When η increases, the intertemporal hedging demand first increases in absolute value, peaks when η is between 2 and 3, and then decreases and approaches zero when η increases further. The investor becomes more and more risk averse and eventually holds only the risk free bond. Even for high values of risk aversion, such as $\eta = 30$, intertemporal hedging demands still exhibit period patterns similar to Figure 22.

7 Conclusion

We introduce a novel class of quadratic term structure models for variance swaps, which are among the most important volatility derivatives. The multivariate state variable follows a quadratic diffusion process. The variance swap curve is quadratic in the state variable, and available in closed form in terms of a linear ordinary differential equation, greatly facilitating empirical applications. Various goodness-of-fit tests show that quadratic models fit variance swap rates remarkably well and largely outperform nested specifications, including popular affine models. We also study dynamic optimal portfolios in variance swaps, stock index and risk free bond. Optimal portfolio weights are available in quasi-closed form in terms of a Taylor series expansion involving conditional moments of the state process, which are available in closed form. The empirical analysis of optimal portfolios shows that optimal portfolio weights in variance swaps follow a short-long strategy, with a short position in long term variance swaps (to earn the negative variance risk premium) and a long position in short term variance swaps (to hedge volatility increases). Such portfolio weights exhibit strong periodic patterns, which depend on the roll-over period and maturity of the variance swaps, and which are mainly borne by the intertemporal hedging demand. The optimal investment in variance swaps can be used either to achieve stable wealth growth or to seek additional risk premium, depending on the risk profile of

[57] Liu (2007) discusses this point for optimal portfolios in stock and bond.

the investor. A Monte Carlo study shows that in both cases the added economic value of variance swaps, in terms of expected utility, is substantial.

τ	Mean	Std	Skew	Kurt	Mean	Std	Skew	Kurt
		In-sample				Out-of-sample		
				Panel A: Variance swap rates				
2	20.76	6.80	0.87	4.09	27.55	11.05	1.39	4.49
3	20.90	6.54	0.78	3.87	27.78	10.21	1.21	3.93
6	21.48	6.32	0.78	3.93	27.94	9.14	1.00	3.51
12	22.25	6.06	0.62	3.19	27.66	8.12	0.72	3.08
24	22.86	5.90	0.55	2.75	27.71	7.03	0.27	2.71
				Panel B: Realized variances				
2	16.42	6.38	0.86	3.20	26.42	14.70	1.80	5.45
3	16.53	6.08	0.69	2.82	26.92	13.89	1.53	4.08
6	16.67	5.79	0.51	2.42	27.26	13.14	1.30	3.22
12	17.02	5.20	0.07	1.77	27.87	11.12	0.60	1.60
24	18.04	5.28	0.22	2.92	28.21	7.52	−0.35	1.29
				Panel C: Realized variance swap payoffs				
2	−1.67	2.80	−1.49	12.27	0.34	11.38	2.71	10.99
3	−1.69	2.79	−1.26	10.34	0.42	10.77	1.96	6.59
6	−1.90	2.77	−1.53	11.55	0.51	10.48	1.40	4.12
12	−2.15	2.54	−1.73	9.95	0.69	9.11	0.32	1.85
24	−2.04	3.10	0.36	6.75	0.71	6.87	−0.11	1.42

Table 12: Dataset summary statistics. Mean, standard deviation (Std), skewness (Skew) and kurtosis (Kurt) of variance swap rates in Panel A, realized variances in Panel B, and realized variance swap payoffs on the S&P 500 index in Panel C. Variance swap rates and realized variances are in volatility percentage units, i.e., $\sqrt{\mathrm{VS}(t,T)} \times 100$ and $\sqrt{\mathrm{RV}(t,T)} \times 100$, respectively. Variance swap payoffs are $(\mathrm{RV}(t,T) - \mathrm{VS}(t,T)) \times 100$. Term τ is in months. In-sample period is from January 4, 1996 to April 2, 2007, and includes 2,832 daily observations. Out-of-sample period is from April 3, 2007 to June 7, 2010, and includes 794 daily observations.

	Univ. quad.		Univ. poly.		Biv. quad.	
Parameter	Est.	S.E.	Est.	S.E.	Est.	S.E.
b_1	2.005	32.240	0.577	2.984		
β_{11}	−0.742	25.440	−0.450	7.594	−5.172	1.439
β_{12}					4.232	0.216
a_1			1	—	1	—
α_1	1	—				
A_1	0.402	0.897	$1.06.10^{-4}$	1.930	3.389	0.266
b_2					0.182	0.025
β_{22}					−0.248	0.002
a_2						
α_2					1	—
A_2					0.010	0.006
λ_0	0.023	1.641	0.007	0.321	−0.028	0.135
λ_1	0.243	25.070	−0.002	0.154	−0.177	1.594
ϕ	0.016	0.017	0.009	0.583	0.017	0.001
ψ	−0.002	0.011	0.003	1.559	0.019	0.001
π	0.002	0.044	0.017	1.784	0.013	0.001
p_3			−0.001	4.068		
p_4			−0.006	2.088		
p_5			0.002	0.349		
Log-likelihood	48,681		48,933		62,671	
AIC	−97,346		−97,844		−125,310	
BIC	−97,298		−97,778		−125,245	

Table 13: Model estimates. Entries are parameter estimates (Est.) for the univariate quadratic, univariate polynomial and bivariate quadratic models and corresponding robust standard errors (S.E.). Identifiable, thus restricted, versions of the following model are estimated: dynamics $dX_{1t} = (b_1 + \beta_{11}X_{1t} + \beta_{12}X_{2t})\,dt + \sqrt{a_1 + \alpha_1 X_{1t} + A_1 X_{1t}^2}\,dW_{1t}$, $dX_{2t} = (b_2 + \beta_{22}X_{2t})\,dt + \sqrt{a_2 + \alpha_2 X_{2t} + A_2 X_{2t}^2}\,dW_{2t}$; spot variance $v_t = \phi + \pi X_{1t} + \psi X_{1t}^2 + p_3 X_{1t}^3 + p_4 X_{1t}^4 + p_5 X_{1t}^5$; market price of risk $(\lambda_0 + \lambda_1 X_{1t})/\sqrt{a_1 + \alpha_1 X_{1t} + A_1 X_{1t}^2}$ for the Brownian motion W_{1t}. An empty entry means that the parameter is set to zero because of model identification. AIC and BIC are Akaike and Bayesian Information Criteria, respectively. Sample data are variance swap rates on the S&P 500, with terms of 2, 3, 6, 12, 24 months, from January 4, 1996 to April 2, 2007, and include 2,832 daily observations.

τ	Bias	RMSE	DM_u	Bias	RMSE	DM_p	Bias	RMSE
				Panel A: In-sample				
	Univ. quad.			Univ. poly.			Biv. quad.	
2	0.10	1.69	9.34	0.13	1.67	9.03	−0.01	0.49
3	0.11	1.14	9.36	0.16	1.09	9.22	0.12	0.40
6	−0.08	0.57	5.88	0.01	0.56	7.03	0.08	0.44
12	−0.17	1.13	6.18	−0.10	1.14	5.86	−0.08	0.29
24	0.21	1.55	6.22	0.03	1.46	5.48	0.07	0.38
				Panel B: Out-of-sample				
	Univ. quad.			Univ. poly.			Biv. quad.	
2	−0.12	2.42	6.30	−0.08	2.14	5.98	0.13	0.80
3	−0.26	1.48	6.06	−0.16	1.33	5.20	−0.03	0.61
6	−0.19	1.35	0.39	0.02	1.41	0.29	−0.13	1.36
12	0.42	1.96	0.13	0.60	1.91	0.06	0.12	1.96
24	0.68	4.08	4.15	0.44	3.67	4.31	0.20	2.26

Table 14: Variance swap pricing errors. The pricing error is defined as the model-based minus observed variance swap rate, both in volatility percentage units, i.e., $(\sqrt{G(\tau, X_t)/\tau} - \sqrt{\mathrm{VS}(t, t+\tau)}) \times 100$. Entries are mean (Bias) and root mean square error (RMSE) of pricing errors for variance swap rates under the univariate quadratic, univariate polynomial and bivariate quadratic models. DM_u (respectively, DM_p) is the Diebold–Mariano test statistic of the univariate quadratic (respectively, polynomial) model versus the bivariate quadratic model, Section 4.3. Under the null hypothesis that the univariate quadratic (respectively, polynomial) model and the bivariate quadratic model have pricing errors of equal magnitude, the DM test statistic is a standard normal. A positive value means that the bivariate quadratic model outperforms the competing univariate model. Term τ is in months. Panel A shows pricing error statistics for the in-sample period, used to estimate the models, which is from January 4, 1996 to April 2, 2007, and includes 2,832 daily observations. Panel B shows pricing error statistics for the out-of-sample period, which is from April 3, 2007 to June 7, 2010, and includes 794 daily observations.

	Martingale		Univ. quad.		Univ. poly.		Biv. quad.	
τ	γ_0	γ_1	γ_0	γ_1	γ_0	γ_1	γ_0	γ_1
Panel A: 1-day ahead prediction								
2	0.14	0.98	2.11	−1.07	−0.36	1.07	0.09	0.98
	0.00	0.00	0.00	0.00	0.01	0.04	0.14	0.22
3	0.09	0.98	2.83	−1.33	−0.06	1.00	0.12	0.97
	0.00	0.01	0.00	0.00	0.29	0.75	0.02	0.01
6	0.06	0.99	15.55	−9.24	0.26	0.95	−0.08	1.01
	0.01	0.04	0.00	0.00	0.00	0.00	0.26	0.37
12	0.04	0.99	0.56	0.90	0.54	0.89	−0.03	1.01
	0.05	0.13	0.00	0.01	0.00	0.00	0.78	0.70
24	0.03	1.01	2.21	0.33	0.51	0.91	0.10	0.98
	0.05	0.14	0.00	0.00	0.16	0.20	0.59	0.63
Panel B: 10-day ahead prediction								
2	0.68	0.88	2.70	−1.16	0.35	0.89	0.38	0.93
	0.00	0.00	0.00	0.00	0.17	0.06	0.06	0.18
3	0.53	0.91	3.49	−1.55	0.49	0.86	0.30	0.94
	0.00	0.01	0.00	0.00	0.00	0.00	0.03	0.07
6	0.42	0.93	−51.00	34.08	0.66	0.84	0.07	0.99
	0.00	0.02	0.00	0.00	0.00	0.00	0.65	0.78
12	0.33	0.95	0.89	0.79	0.83	0.83	0.10	0.99
	0.00	0.03	0.00	0.00	0.00	0.00	0.54	0.74
24	0.30	0.96	2.47	0.30	0.88	0.84	0.34	0.94
	0.00	0.02	0.00	0.00	0.03	0.03	0.11	0.22

Table 15: Variance swap predictive regressions. For each model and term, entries report time series regressions of future actual variance swap rates on a constant and a d-day ahead model-based prediction, i.e., $\text{VS}(t, t+\tau) = \gamma_0 + \gamma_1 \, \mathbb{E}_{\mathbb{P}}[G(\tau, X_t)/\tau | \mathcal{F}_{t-d}] + \text{error}_t$, where d is either 1-day (Panel A) or 10-day (Panel B), and $\mathbb{E}_{\mathbb{P}}[G(\tau, X_t)/\tau | \mathcal{F}_{t-d}]$ is the time $t-d$ model-based, conditional prediction of the τ-variance swap rate observed at time t. Variance swap rates are in variance percentage units, i.e., $\text{VS}(t, t+\tau) \times 100$. For each term τ, the first row reports estimates of γ_0 and γ_1, the second row reports the p-value of the null hypotheses $H_0 : \gamma_0 = 0$ and $H_0 : \gamma_1 = 1$, respectively. If model-based variance swap rate predictions are unbiased, then $\gamma_0 = 0$. If model-based variance swap rate predictions are efficient, then $\gamma_1 = 1$. Robust standard errors are computed using Newey and West (1987) covariance matrix estimator with the number of lags optimally chosen according to Andrews (1991). The martingale model is a benchmark model in which the future actual $\text{VS}(t, t+\tau)$ is predicted using the past actual $\text{VS}(t-d, t-d+\tau)$. Term τ is in months. The sample period is from January 4, 1996 to June 7, 2010, and includes 3,626 daily observations.

Panel A: Risk aversion $\eta = 5$

	C.E.	*Rate*	*S.R.*	$\mathbb{E}_{\mathbb{P}}[V_T]$	*Rebalance*
Optimal portfolio	101.06	2.53	1.59	106.08	daily
Proxy portfolio	100.95	2.47	1.47	106.08	monthly, yearly
Stock index	90.98	−2.73	1.18	110.42	—

Panel B: Risk aversion $\eta = 1$

	C.E.	*Rate*	*S.R.*	$\mathbb{E}_{\mathbb{P}}[V_T]$	*Rebalance*
Optimal portfolio	104.79	4.34	1.44	114.06	daily
Proxy portfolio	104.42	4.16	1.51	114.42	monthly, yearly
Stock-bond portfolio	103.52	3.73	1.39	110.99	daily
Stock index	103.47	3.70	1.18	110.42	—

Table 16: Monte Carlo study of optimal portfolios. Results are based on 10,000 simulated trajectories of bivariate state process X_t (Section 5.3), variance swap rates, and stock index. Optimal portfolio: wealth is optimally invested in the risk free bond, stock index, 3-month and 2-year variance swaps. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. Proxy portfolio: wealth is invested as in the optimal portfolio, but positions are rebalanced less frequently, namely stock index and 3-month variance swap positions are rebalanced monthly, 2-year variance swap position is rebalanced yearly. Stock-bond portfolio: for the risk aversion $\eta = 1$, wealth is optimally invested in the risk free bond and stock index. Stock index: initial wealth is invested in the stock index and positions are not rebalanced. Initial wealth is 100. *C.E.* is the certainty equivalent defined as $\mathrm{e}^{-rT}\mathcal{C}$, and $\mathcal{C}$ is such that $u(\mathcal{C}) = \mathbb{E}_{\mathbb{P}}[u(V_T)]$, where $u(V) = V^{1-\eta}/(1-\eta)$ for $\eta \neq 1$ and $u(V) = \log(V)$ for $\eta = 1$, T is the investment horizon of 2 years, r is the risk free rate set to 2%, and V_T is the terminal wealth. *Rate* is the percentage implied annual rate of return, i.e., $\log(\mathcal{C}/100)/T \times 100$. *S.R.* is the Sharpe ratio of the corresponding portfolio. $\mathbb{E}_{\mathbb{P}}[V_T]$ is the average terminal wealth, obtained by averaging terminal wealths across sample paths. Panel A is for the risk aversion $\eta = 5$. Panel B is for the risk aversion $\eta = 1$, i.e., the logarithmic utility case.

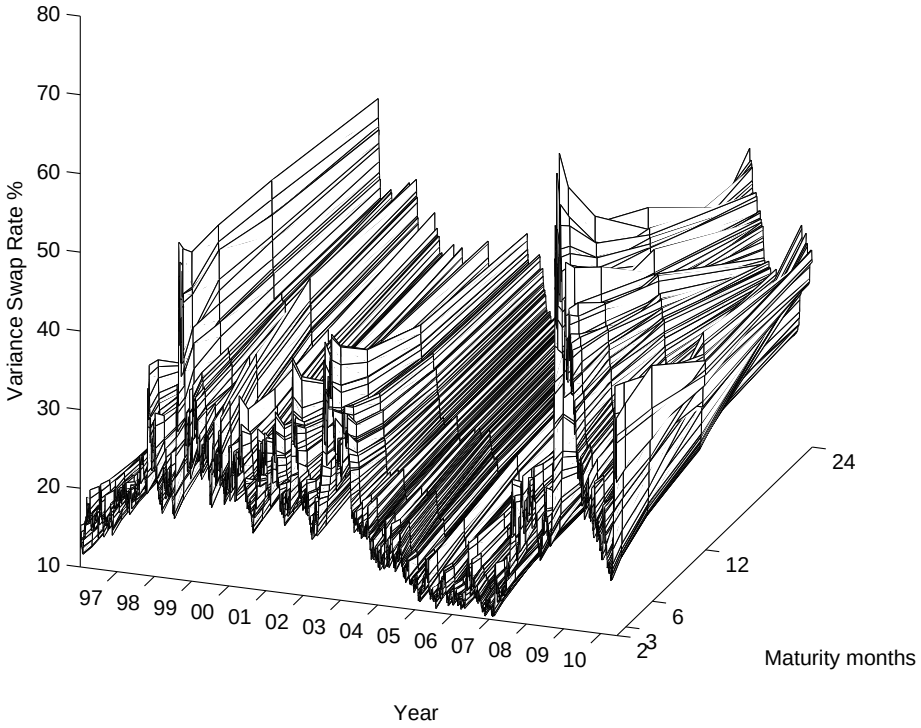


Figure 16: Term structure of variance swaps rates. Variance swap rates on the S&P 500 in volatility percentage units, $\sqrt{\mathrm{VS}(t,T)} \times 100$. Terms are 2, 3, 6, 12, 24 months. Sample period is from January 4, 1996 to June 7, 2010.

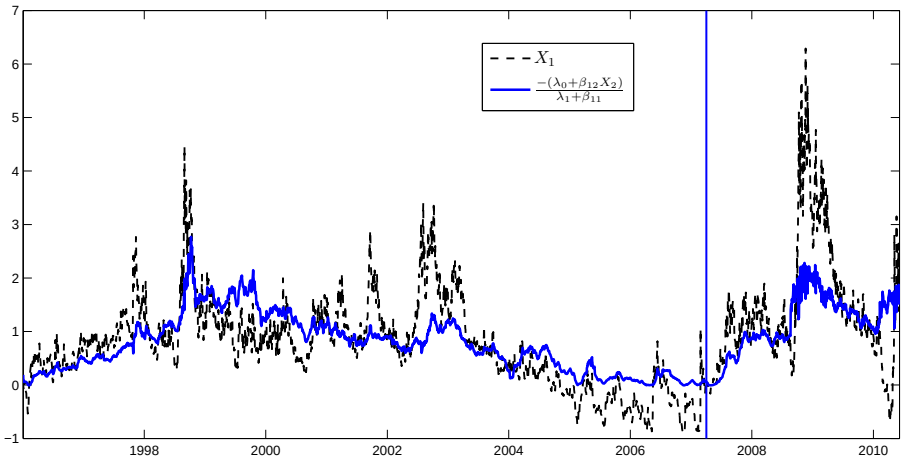


Figure 17: Time series evolution of state process. In the bivariate quadratic model in Section 3.2, X_{1t} is in Class 1 and X_{2t} is in Class 3; Proposition 3.2. The model is fitted to daily variance swap rates on the S&P 500, from January 4, 1996 to April 2, 2007, and terms of 2, 3, 6, 12, 24 months. The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.

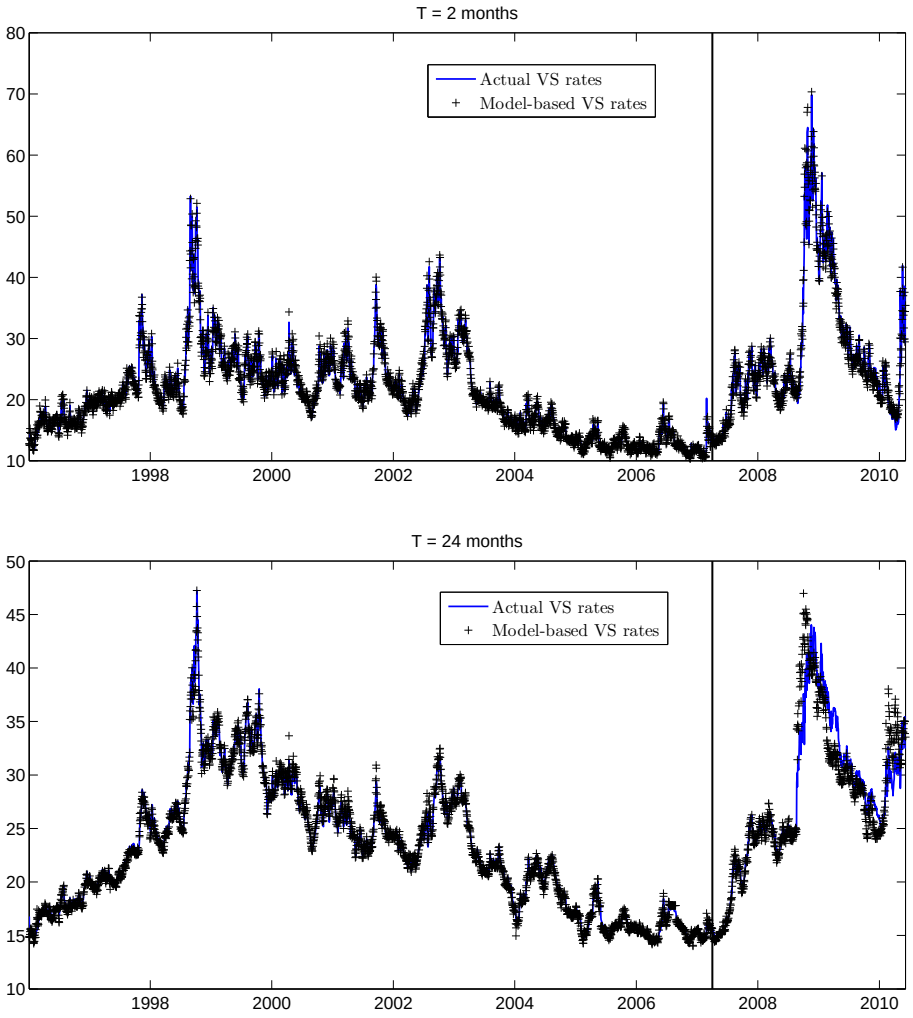


Figure 18: Actual and model-based variance swap rates. Model-based variance swap rates are from the bivariate quadratic model in Section 3.2, with X_{1t} in Class 1 and X_{2t} in Class 3; Proposition 3.2. The model is fitted to daily variance swap rates on the S&P 500, from January 4, 1996 to April 2, 2007, and terms of 2, 3, 6, 12, 24 months. Variance swap rates are in volatility percentage units, i.e., $\sqrt{\mathrm{VS}(t,T)} \times 100$. Upper graph: variance swap rates with 2-month term (shortest term in our sample). Lower graph: variance swap rates with 2-year term (longest term in our sample). The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.

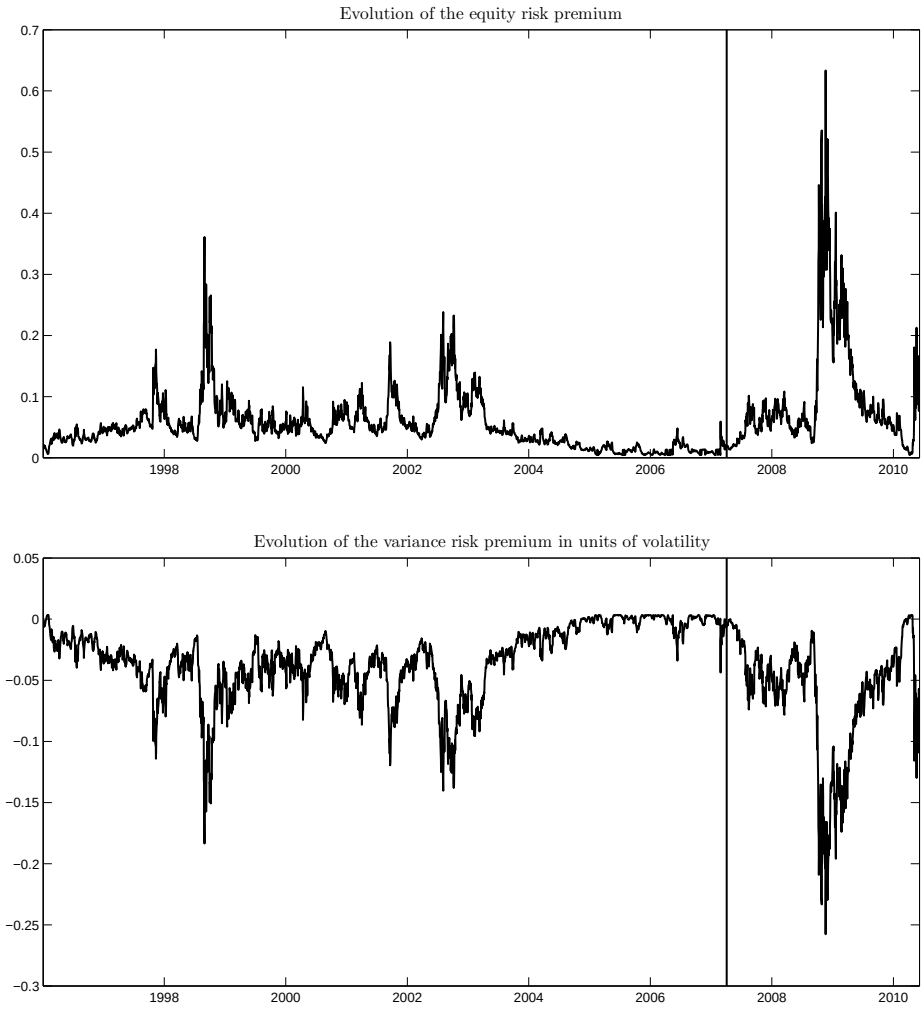


Figure 19: Equity and variance risk premium. Upper graph: equity risk premium defined as $(\mathbb{E}_{\mathbb{P}}[dS_t/S_t|\mathcal{F}_t] - \mathbb{E}_{\mathbb{Q}}[dS_t/S_t|\mathcal{F}_t])/dt$, where S_t is the stock index. Lower graph: variance risk premium, in units of volatility, defined as $(\mathbb{E}_{\mathbb{P}}[dv_t|\mathcal{F}_t] - \mathbb{E}_{\mathbb{Q}}[dv_t|\mathcal{F}_t])/(dt\sqrt{v_t})$, where v_t is the spot variance. Both risk premiums are derived from the bivariate quadratic model in Section 5.3. The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.

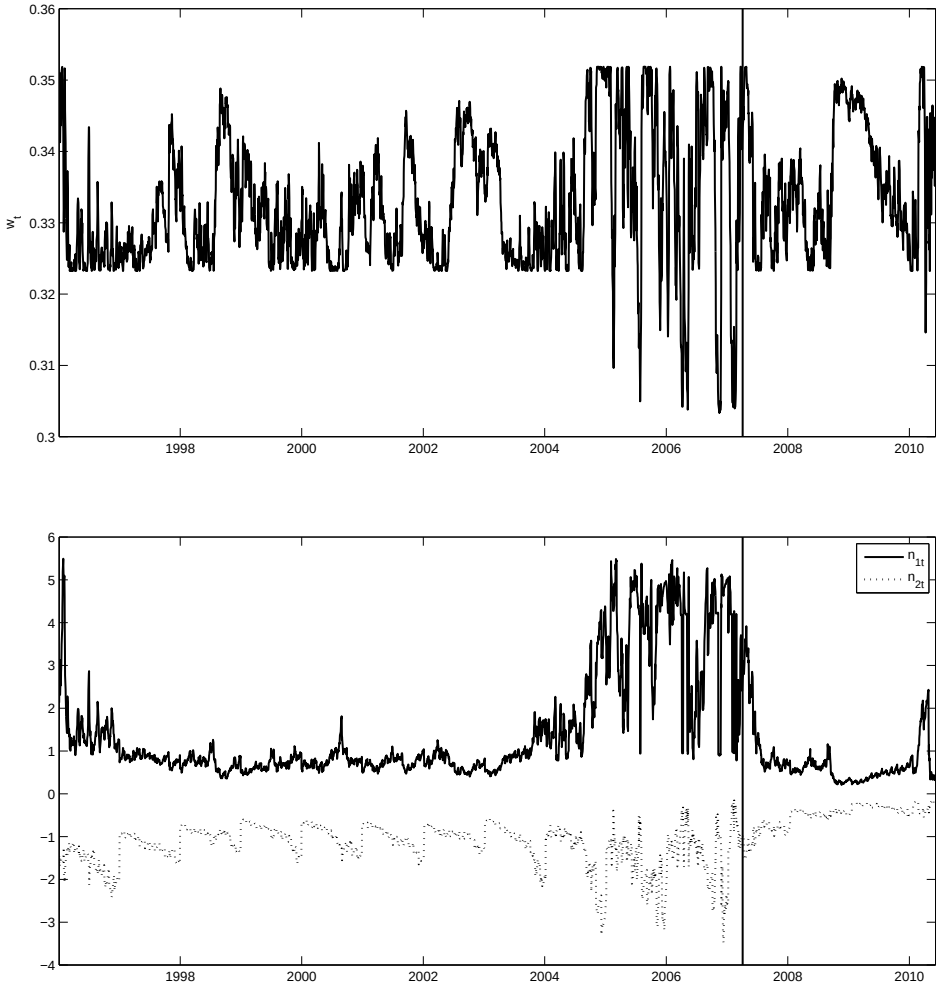


Figure 20: Optimal portfolio. Wealth is optimally invested in the risk free bond, stock index, 3-month and 2-year variance swaps. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. The risk aversion is $\eta = 5$. n_{1t} is the optimal fraction of wealth invested in the 3-month variance swap. n_{2t} is the optimal fraction of wealth invested in the 2-year variance swap. Upper graph: optimal portfolio weight in the stock index, w_t. Lower graph: optimal portfolio weights in variance swaps, n_{1t} and n_{2t}. The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.

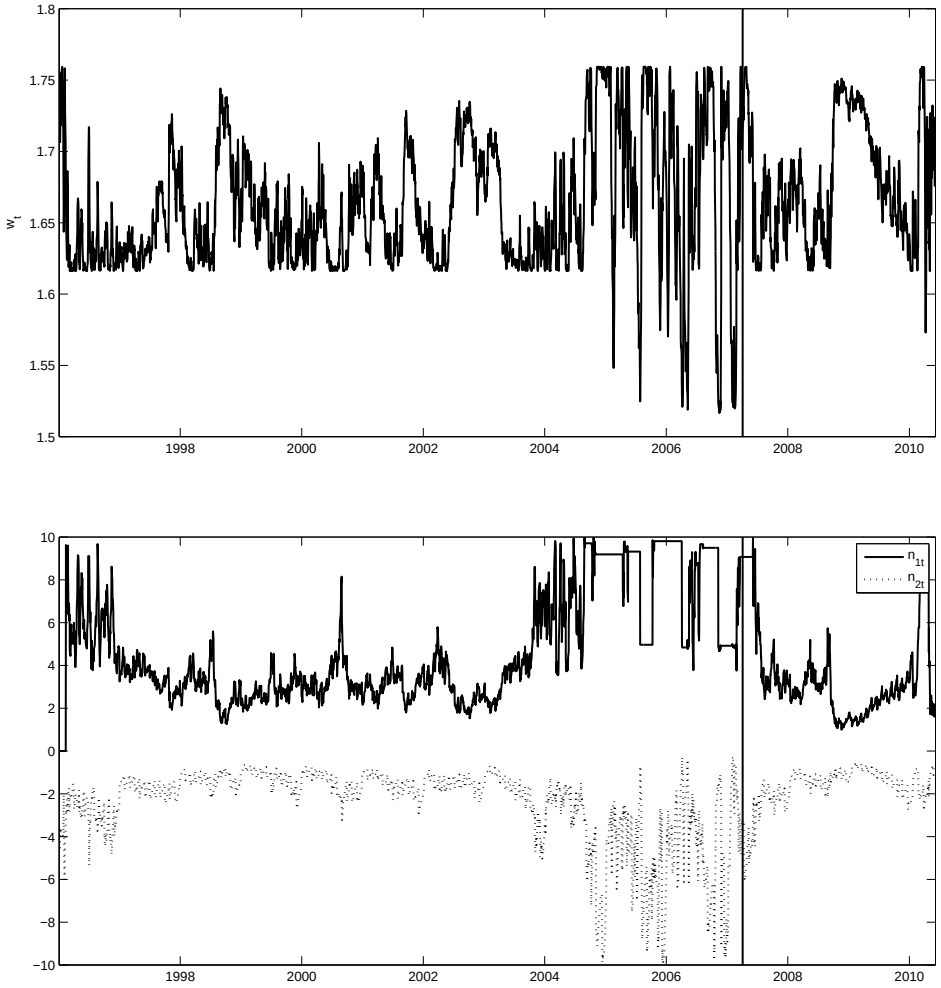


Figure 21: Optimal portfolio for log-investor. Wealth is optimally invested in the risk free bond, stock index, 3-month and 2-year variance swaps. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. The risk aversion is $\eta = 1$. n_{1t} is the optimal fraction of wealth invested in the 3-month variance swap. n_{2t} is the optimal fraction of wealth invested in the 2-year variance swap. Upper graph: optimal portfolio weight in the stock index, w_t. Lower graph: optimal portfolio weights in variance swaps, n_{1t} and n_{2t}. The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.

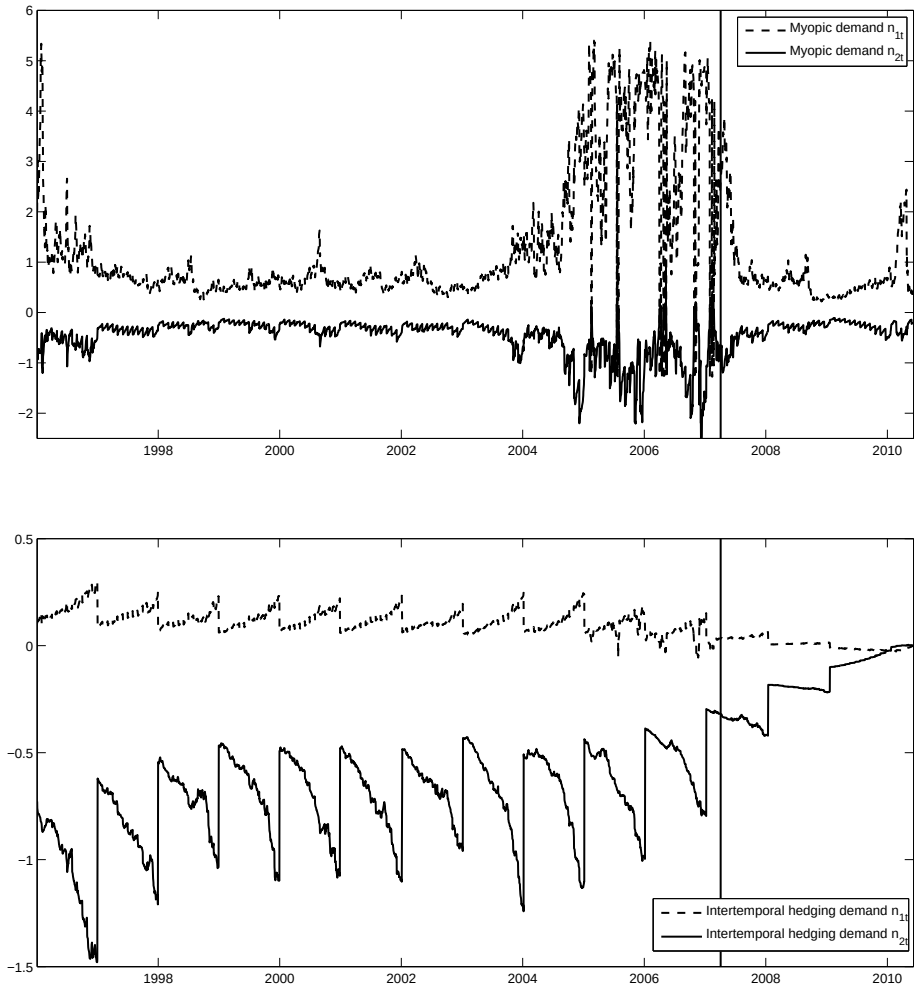


Figure 22: Myopic and intertemporal hedging demands. Wealth is optimally invested in the risk free bond, stock index, 3-month and 2-year variance swaps. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. The risk aversion is $\eta = 5$. n_{1t} is the optimal fraction of wealth invested in the 3-month variance swap. n_{2t} is the optimal fraction of wealth invested in the 2-year variance swap. Upper graph: myopic component. Lower graph: intertemporal hedging demand. The myopic component and intertemporal hedging demand in the risky assets are in (5.7). The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.

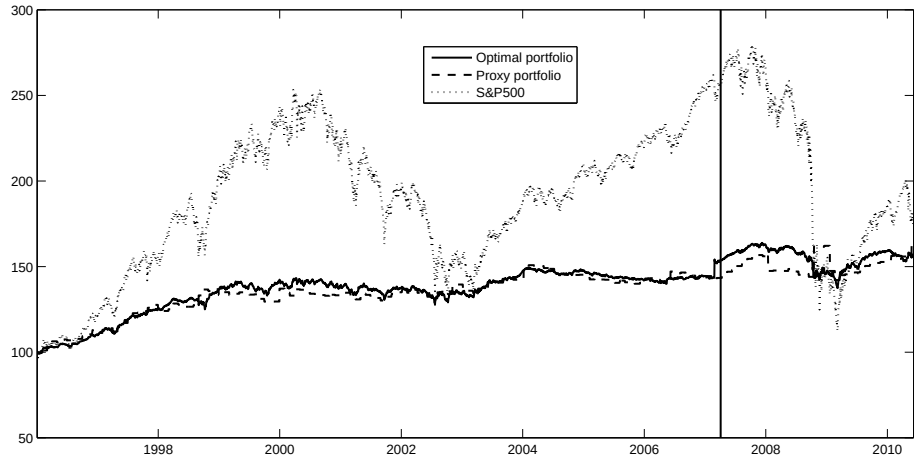


Figure 23: Wealth process. Wealth is optimally invested in the risk free bond, stock index, 3-month and 2-year variance swaps. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. Proxy portfolio is rebalanced less frequently: stock index and 3-month variance swap positions are rebalanced monthly, 2-year variance swap position is rebalanced yearly. The risk aversion is $\eta = 5$. S&P 500 is normalized to 100. The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.

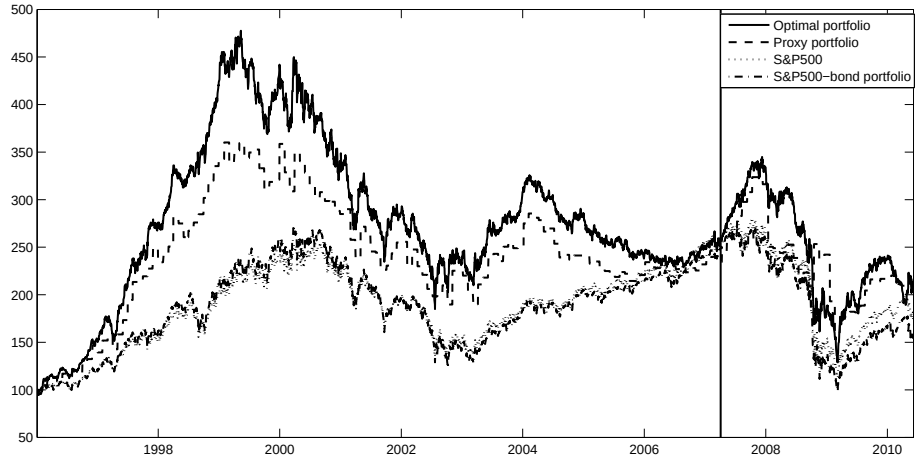


Figure 24: Wealth process for log-investor. Wealth is optimally invested in the risk free bond, stock index, 3-month and 2-year variance swaps. Variance swaps are rolled over monthly and yearly, respectively. Optimal portfolio is rebalanced daily. Proxy portfolio is rebalanced less frequently: stock index and 3-month variance swap positions are rebalanced monthly, 2-year variance swap position is rebalanced yearly. "S&P500-bond portfolio" optimally invests in risk free bond and stock index, and is rebalanced daily. The risk aversion is $\eta = 1$. S&P 500 is normalized to 100. The vertical line is April 3, 2007, i.e., beginning of the out-of-sample period.

ONLINE APPENDIX TO
Quadratic Variance Swap Models

This appendix provides technical derivation and proofs.

1 Kolmogorov Backward Equation

This section provides some technical results on diffusion processes which form the background of several proofs in this paper. As at the beginning of Section 3, let X_t be a diffusion process taking values in some state space $\mathcal{X} \subset \mathbb{R}^m$ and satisfying the SDE (3.1) where W_t is a standard d-dimensional Brownian motion under the risk neutral measure $\mathbb{Q}$. The following assumption is obviously met by all quadratic processes in this paper.

Assumption 1.1. *The drift and dispersion functions $\mu(x)$ and $\Sigma(x)$ are assumed to be continuous maps from $\mathcal{X}$ to $\mathbb{R}^m$ and $\mathbb{R}^{m\times m}$ satisfying the linear growth condition*

$$\|\mu(x)\|^2 + \|\Sigma(x)\|^2 \le K(1 + \|x\|^2), \quad x \in \mathcal{X} \tag{1.1}$$

for some finite constant K.

Lemma 1.1. *Let $g(x)$ be some C^2-function on $\mathcal{X}$, and suppose $f(\tau, x)$ is a $C^{1,2}$-function on $[0, +\infty)\times\mathcal{X}$ whose x-gradient satisfies a polynomial growth condition*

$$\|\nabla_x f(\tau, x)\| \le K(1 + \|x\|^p), \quad \tau \le T, \quad x \in \mathcal{X} \tag{1.2}$$

for some finite constant $K = K(T)$ and some $p \ge 1$, for all finite T.

If $f(\tau, x)$ satisfies the Kolmogorov backward equation

$$\begin{aligned} \frac{\partial f(\tau, x)}{\partial \tau} &= \sum_{i=1}^m \mu_i(x)\frac{\partial f(\tau, x)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^m \left(\Sigma(x)\Sigma(x)^\top\right)_{ij} \frac{\partial^2 f(\tau, x)}{\partial x_i \partial x_j} \\ f(0, x) &= g(x) \end{aligned} \tag{1.3}$$

for all $\tau \ge 0$ and $x \in \mathcal{X}$, then

$$f(T - t, X_t) = \mathbb{E}_{\mathbb{Q}}[g(X_T) \mid \mathcal{F}_t] \quad \text{for all } t \le T < \infty. \tag{1.4}$$

Proof. Fix some finite T. Itô's formula applied to $M_t = f(T - t, X_t)$ gives

$$dM_t = D_t\, dt + \nabla_x f(T - t, X_t)\, \Sigma(X_t)\, dW_t$$

with drift term

$$D_t = -\frac{\partial f(T-t,X_t)}{\partial \tau} + \sum_{i=1}^{m} \mu_i(X_t) \frac{\partial f(T-t,X_t)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{m} \left(\Sigma(X_t)\Sigma(X_t)^\top \right)_{ij} \frac{\partial^2 f(T-t,X_t)}{\partial x_i \partial x_j} \quad (1.5)$$

which vanishes by assumption. Hence M_t is a $\mathbb{Q}$-local martingale with $M_T = g(X_T)$. It remains to be shown that M_t is a true $\mathbb{Q}$-martingale. Assumption (1.2) implies

$$\begin{aligned} \mathbb{E}\left[\int_0^T \left\| \nabla_x f(T-s,X_s)\, \Sigma(X_s) \right\|^2 ds \right] &\le \mathbb{E}\left[\int_0^T \left\| \nabla_x f(T-s,X_s) \right\|^2 \left\| \Sigma(X_s) \right\|^2 ds \right] \\ &\le K \left(1 + \mathbb{E}\left[\sup_{s \le T} \|X_s\|^{2p} \right] \right) \end{aligned}$$

for some finite constant K. Lemma 1.2 below now yields the assertion. □

The following useful lemma follows from Karatzas and Shreve (1991, Problem V.3.15). For the convenience of the reader we provide a self-contained short proof.

Lemma 1.2. *The above diffusion process X_t with $X_0 = x \in \mathcal{X}$ satisfies $\mathbb{E}\left[\sup_{s \le T} \|X_s\|^{2p}\right] < \infty$, for all $p \ge 1$ and finite T.*

Proof. Let $n \ge 1$ and define the finite stopping time $T_n = \inf\{t \mid \|X_t\| \ge n\}$. The stopped process $X_t^{T_n} = X_{t \wedge T_n}$ satisfies

$$X_t^{T_n} = x + \int_0^t \mu(X_s^{T_n}) 1_{\{s \le T_n\}}\, ds + \int_0^t \Sigma(X_s^{T_n}) 1_{\{s \le T_n\}}\, dW_s =: x + D_t + M_t.$$

We fix a finite T. In what follows, $K_1, K_2, \ldots$ denote some universal finite constants, which only depend on T. First, observe that the linear growth condition (1.1) implies the pathwise inequality

$$\sup_{s \le t} \|D_s\|^{2p} \le K_1 \int_0^t \|\mu(X_u^{T_n})\|^{2p} du \le K_2 \int_0^t \left(1 + \sup_{s \le u} \|X_s^{T_n}\|^{2p} \right) du.$$

Next, the Burkholder–Davis–Gundy inequality, Karatzas and Shreve (1991, Theorem III.3.28), applied to the continuous local martingale M_t, combined with (1.1), yields

$$\mathbb{E}\left[\sup_{s \le t} \|M_s\|^{2p} \right] \le K_3 \int_0^t \mathbb{E}\left[\left\| \Sigma(X_u^{T_n}) \right\|^{2p} \right] du \le K_4 \int_0^t \left(1 + \mathbb{E}\left[\sup_{s \le u} \left\| X_s^{T_n} \right\|^{2p} \right] \right) du.$$

Combining these inequalities, we obtain

$$\mathbb{E}\left[\sup_{s\le t}\left\|X_s^{T_n}\right\|^{2p}\right] \le K_5\left(x^{2p}+t+\int_0^t \mathbb{E}\left[\sup_{s\le u}\left\|X_s^{T_n}\right\|^{2p}\right]du\right).$$

By dominated convergence, the nonnegative function $[0,T]\ni t\mapsto \mathbb{E}\left[\sup_{s\le t}\|X_s^{T_n}\|^{2p}\right]$ is continuous. Applying Gronwall's inequality, Karatzas and Shreve (1991, Problem V.2.7), to it yields

$$\mathbb{E}\left[\sup_{s\le T}\left\|X_s^{T_n}\right\|^{2p}\right] \le K_5\left(x^{2p}+T+\int_0^T \left(x^{2p}+u\right)K_5\mathrm{e}^{K_5(T-u)}du\right).$$

The right hand side does not depend on n. Letting $n\to\infty$, monotone convergence thus proves the claim. □

2 X_t is Necessarily Quadratic

The aim of this section is to show that, under some mild technical conditions, a quadratic term structure of variance swap rates implies that the state process X_t be quadratic. In addition to Assumption 1.1 in Appendix 1, we assume the following:

Assumption 2.1. *The SDE* (3.1) *is* well posed *in* $\mathcal{X}$*. That is, for any* $x\in\mathcal{X}$ *there exists a* $\mathcal{X}$*-valued weak solution* $X=X^x$ *of* (3.1) *with* $X_0=x$ *which is unique in law. We let* X_t *be realized on the canonical space of continuous paths* $\omega:[0,\infty)\to\mathcal{X}$*. It is well known that in this case* X_t *has the strong Markov property, e.g., Karatzas and Shreve (1991, Chapter V).*

Assumption 2.2. *The spot variance is given by* $v_t=g(X_t)$ *for some* C^2*-function* $g(x)$ *on* $\mathcal{X}$*.*

Assumption 2.3. *The law* $\mathbb{Q}=\mathbb{Q}_x$ *of the state process* $X=X^x$ *is risk neutral for any initial state* $X_0=x\in\mathcal{X}$*, and the variance swap curve is given by* $\mathrm{VS}(t,T)=\frac{1}{T-t}\int_t^T \mathbb{E}_{\mathbb{Q}}[v_s\mid X_t]\,ds$*.*

Hence $\mathrm{VS}(t,T)$ is a function of the prevailing state X_t and term $T-t$. It is well known that, under suitable regularity conditions, this function can be characterized by the Kolmogorov backward equation. The following lemma makes this explicit.

Lemma 2.1. *Suppose* $f(\tau,x)$ *is a* $C^{1,2}$*-function on* $[0,+\infty)\times\mathcal{X}$ *whose* x*-gradient satisfies a polynomial growth condition* (1.2) *for some finite constant* $K=K(T)$ *and some* $p\ge 1$*, for all finite* T*. Then, under the above assumptions, the converse of Lemma 1.1 holds true: validity of* (1.4) *for all initial states* $X_0=x\in\mathcal{X}$ *implies that* $f(\tau,x)$ *satisfies the Kolmogorov backward equation* (1.3)*.*

Proof. By assumption, $M_t=f(T-t,X_t)$ is a $\mathbb{Q}$-martingale. Hence its drift, given in (1.5), has to vanish a.s. for all $t\le T<\infty$ and for all initial states $x\in\mathcal{X}$. This is equivalent to (1.3). □

We are ready to state and prove the converse of Proposition 3.1.

Proposition 2.1. *Assume that the variance swap model admits a quadratic term structure. That is, $G(\tau, x)$ in (3.5) is a quadratic function in x, $G(\tau, x) = \Phi(\tau)+\Psi(\tau)^\top x+x^\top \Pi(\tau)\, x$, for some C^2-functions $\Phi : [0, +\infty) \to \mathbb{R}$, $\Psi : [0, +\infty) \to \mathbb{R}^m$, and $\Pi : [0, +\infty) \to \mathbb{S}^m$. Then the spot variance function is quadratic, $g(x) = \phi + \psi^\top x + x^\top \pi\, x$, with parameters given by $\phi = d\Phi(0)/d\tau$, $\psi = d\Psi(0)/d\tau$, and $\pi = d\Pi(0)/d\tau$. Moreover, the following holds:*

(i) *Suppose $\Psi_i(\tau)$ and $\Pi_{ij}(\tau)$, $1 \le i \le j \le m$, are linearly independent functions. Assume the state space $\mathcal{X}$ contains $\{\lambda x \mid x \in O, \lambda \ge 1\}$ for some open set O in $\mathbb{R}^m$. Then the process X_t is quadratic with drift and diffusion functions of the form (3.2)–(3.3). The functions $\Phi(\tau)$, $\Psi(\tau)$, and $\Pi(\tau)$ satisfy the linear ordinary differential equations (3.6).*

(ii) *If $\Pi(\tau) \equiv 0$, and if $\Psi_i(\tau)$, $1 \le i \le m$, are linearly independent functions, then the drift function of the state process X_t is affine of the form (3.2). The functions $\Phi(\tau)$ and $\Psi(\tau)$ satisfy the linear ordinary differential equations*

$$\begin{aligned} \frac{d\Phi(\tau)}{d\tau} &= \phi + b^\top \Psi(\tau), & \Phi(0) &= 0 \\ \frac{d\Psi(\tau)}{d\tau} &= \psi + \beta^\top \Psi(\tau), & \Psi(0) &= 0. \end{aligned}$$

Proof. Notice that the assumptions of Lemma 2.1 are satisfied by the function $f(\tau, x) = \partial G(\tau, x)/\partial\tau$. Moreover, note that by assumption, $g(x) = f(0, x) = \phi + \psi^\top x + x^\top \pi\, x$ for $\phi = d\Phi(0)/d\tau$, $\psi = d\Psi(0)/d\tau$, and $\pi = d\Pi(0)/d\tau$. We denote by $c(x) = \Sigma(x)\Sigma(x)^\top$ the diffusion function of X_t. Integrating the Kolmogorov backward equation (1.3) for $f(\tau, x)$ in τ leads to

$$\begin{aligned} &\frac{d\Phi(\tau)}{d\tau} - \phi + \left(\frac{d\Psi(\tau)}{d\tau} - \psi\right)^\top x + x^\top \left(\frac{d\Pi(\tau)}{d\tau} - \pi\right) x \\ &\quad = \sum_{i=1}^m \Psi_i(\tau)\mu_i(x) + \sum_{i,j=1}^m \Pi_{ij}(\tau)\left(\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x)\right) \\ &\quad = \sum_{i=1}^m \Psi_i(\tau)\mu_i(x) + \sum_{i=1}^m \Pi_{ii}(\tau)\left(2\mu_i(x)x_i + c_{ii}(x)\right) + 2\sum_{i<j} \Pi_{ij}(\tau)\left(\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x)\right) \end{aligned} \tag{2.1}$$

for all τ and $x \in \mathcal{X}$. On the left hand side of this equation there is quadratic polynomial.

If $\Psi_i(\tau)$ and $\Pi_{ij}(\tau)$, $i \le j$, are linearly independent, we obtain that $\mu_i(x)$ and $\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x)$ are polynomials in x of degree less than or equal 2. If, moreover, $\mathcal{X}$ contains $\{\lambda x \mid x \in O, \lambda \ge 1\}$ for some open set O in $\mathbb{R}^m$ then the linear growth condition (1.1) implies that $\mu_i(x)$ is in fact affine in x, that is of the form (3.2). Plugging this in $\mu_i(x)x_j + \mu_j(x)x_i + c_{ij}(x)$ yields (3.3). Plugging these expressions back in (2.1), and separating the powers of x, we arrive at the linear ordinary differential equations (3.6). This proves part (i). Part (ii) follows using a similar argument. □

3 Univariate and Bivariate Quadratic Term Structures

The functions $\Phi(\tau)$, $\Psi(\tau)$, and $\Pi(\tau)$ for the univariate quadratic model in Section 3.1 satisfy the linear ordinary differential equations

$$\begin{aligned}
\frac{d\Phi(\tau)}{d\tau} &= \phi + b\Psi(\tau) + a\Pi(\tau), && \Phi(0) = 0 \\
\frac{d\Psi(\tau)}{d\tau} &= \psi + \beta\Psi(\tau) + (2b + \alpha)\Pi(\tau), && \Psi(0) = 0 \\
\frac{d\Pi(\tau)}{d\tau} &= \pi + (2\beta + A)\Pi(\tau), && \Pi(0) = 0
\end{aligned} \tag{3.1}$$

for real parameters ϕ, ψ, π.

The vector- and matrix-valued functions $\Phi(\tau)$, $\Psi(\tau)$, and $\Pi(\tau)$ for the bivariate quadratic model in Section 3.2 satisfy the linear ordinary differential equations

$$\begin{aligned}
\frac{d\Phi(\tau)}{d\tau} &= \phi + b^\top \Psi(\tau) + a_1\Pi_{11}(\tau) + a_2\Pi_{22}(\tau), && \Phi(0) = 0 \\
\frac{d\Psi(\tau)}{d\tau} &= \begin{pmatrix} \psi \\ 0 \end{pmatrix} + \beta^\top \Psi(\tau) + 2\Pi(\tau)\, b + \begin{pmatrix} \alpha_1\Pi_{11}(\tau) \\ \alpha_2\Pi_{22}(\tau) \end{pmatrix}, && \Psi(0) = 0 \\
\frac{d\Pi(\tau)}{d\tau} &= \begin{pmatrix} \pi & 0 \\ 0 & 0 \end{pmatrix} + \beta^\top \Pi(\tau) + \Pi(\tau)\,\beta + \begin{pmatrix} A_1\Pi_{11}(\tau) & 0 \\ 0 & A_2\Pi_{22}(\tau) \end{pmatrix}, && \Pi(0) = 0
\end{aligned}$$

for real parameters ϕ, ψ, π. For the purpose of solving these ordinary differential equations, it is useful to vectorize them by setting $Q(\tau) = (\Phi(\tau),\, \Psi_1(\tau),\, \Psi_2(\tau),\, \Pi_{11}(\tau),\, \Pi_{12}(\tau),\, \Pi_{22}(\tau))^\top$. The above system then reads (for $\beta_{21} = 0$):

$$\frac{dQ(\tau)}{d\tau} = \begin{pmatrix} \phi \\ \psi \\ 0 \\ \pi \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & b_1 & b_2 & a_1 & 0 & a_2 \\ 0 & \beta_{11} & \beta_{21} & 2b_1 + \alpha_1 & 2b_2 & 0 \\ 0 & \beta_{12} & \beta_{22} & 0 & 2b_1 & 2b_2 + \alpha_2 \\ 0 & 0 & 0 & 2\beta_{11} + A_1 & 2\beta_{21} & 0 \\ 0 & 0 & 0 & \beta_{12} & \beta_{11} + \beta_{22} & \beta_{21} \\ 0 & 0 & 0 & 0 & 2\beta_{12} & 2\beta_{22} + A_2 \end{pmatrix} Q(\tau), \quad Q(0) = 0. \tag{3.2}$$

4 Proof of Proposition 3.2

It follows by inspection that the quadratic property is invariant with respect to affine transformations $\mathcal{X} \to c + \gamma\mathcal{X}$, $x \mapsto c + \gamma x$ of the state variable, for any real parameters c and $\gamma \neq 0$. Indeed, the transformed process $\hat{X}_t = c + \gamma X_t$ is quadratic with drift and diffusion functions

$$\begin{aligned}
\hat{b}(\hat{x}) &= b\gamma - \beta c + \beta\hat{x} \equiv \hat{b} + \hat{\beta}\hat{x} \\
\hat{a}(\hat{x}) &= a\gamma^2 - \alpha\gamma c + Ac^2 + (\alpha\gamma - 2Ac)\,\hat{x} + A\hat{x}^2 \equiv \hat{a} + \hat{\alpha}\hat{x} + \hat{A}\hat{x}^2.
\end{aligned}$$

The discriminant of $\hat{a}(\hat{x})$ satisfies $\hat{D} = \gamma^2 D$. This proves that Classes 1–3 in Proposition 3.2 form equivalence classes with respect to affine transformations of X_t. It remains to be shown that for any class there exists an affine transformation such that the drift and diffusion functions are of the desired form.

Class 1: Assume first that $A > 0$ and $D < 0$. Any affine transformation with $c = \frac{\alpha}{2A}\gamma$ and $\gamma = \pm\sqrt{\frac{4A}{-D}}$ yields $\hat{a}(\hat{x}) = 1 + A\hat{x}^2$. The sign of γ can be chosen such that $\hat{b} = (b - \beta\frac{\alpha}{2A})\gamma$ be nonnegative as desired. Since the diffusion function has no real zeros, the canonical state space is $\hat{\mathcal{X}} = \mathbb{R}$, e.g., Filipović (2009, Lemma 10.11). If $A = \alpha = 0$ and $a > 0$, we set $\gamma = 1/\sqrt{a}$, and note that c can be chosen such that $\hat{b}$ becomes zero.

Class 2: Assume first that $A > 0$ and $D = 0$. Any affine transformation with $c = \frac{\alpha}{2A}\gamma$ yields $\hat{a}(\hat{x}) = A\hat{x}^2$. The factor γ can be chosen such that $\hat{b} = (b - \beta\frac{\alpha}{2A})\gamma$ is either 1 or 0. A standard comparison result for diffusion processes, Karatzas and Shreve (1991, Proposition V.2.18), shows that $\hat{X}_t$ is bounded from below by the positive geometric Brownian motion $dZ_t = \beta Z_t + \sqrt{A} Z_t\, dW_t$. Hence the canonical state space is $\hat{\mathcal{X}} = (0, +\infty)$. If $A = \alpha = a = 0$, we can chose γ and c so that $\hat{b}$ becomes zero.

Class 3: Assume first that $A > 0$ and $D > 0$. Any affine transformation with $c = \frac{\alpha \pm \sqrt{D}}{2A}\gamma$ and $\gamma = \frac{\pm 1}{\sqrt{D}}$ yields $\hat{a}(\hat{x}) = \hat{x} + A\hat{x}^2$. The sign of γ can be chosen such that $\hat{b} = (b - \beta\frac{\alpha \pm \sqrt{D}}{2A})\gamma$ be nonnegative. Standard stochastic invariance results for diffusion processes, e.g., Filipović (2009, Lemma 10.11), then show that $\hat{X}_t \geq 0$ for all t whenever $\hat{X}_0 \geq 0$. We now claim that $\hat{b} \geq \frac{1}{2}$ is necessary and sufficient for the canonical state space $\hat{\mathcal{X}}$ not to contain 0. Indeed, elementary calculations show that the scale function of $\hat{X}_t$ is

$$p(\hat{x}) = \int_1^{\hat{x}} \left(\frac{(1+A)y}{1+Ay}\right)^{-2\hat{b}} \left(\frac{1+Ay}{1+A}\right)^{-\frac{2\hat{\beta}}{A}} dy.$$

It satisfies $p(\hat{x}) = p(r)\mathbb{P}[\tau_r < \tau_R] + p(R)\mathbb{P}[\tau_r > \tau_R]$ for any $0 \leq r < \hat{X}_0 = \hat{x} < R$, and hitting times defined by $\tau_c = \inf\{t \geq 0 \mid \hat{X}_t = c\}$, see Karatzas and Shreve (1991, Section V.5.C). Since $\tau_R \uparrow \infty$ for $R \uparrow \infty$, it follows that $\mathbb{P}[\tau_0 = \infty] = 1$ if and only if $p(0) = -\infty$, e.g., Filipović (2009, Exercise 10.12). The latter is equivalent to $2\hat{b} \geq 1$, which proves the claim. If $A = 0$ and $\alpha \neq 0$, we set $\gamma = 1/\alpha$, and chose c such that $\hat{a}$ becomes zero. Note that the conditions on $\hat{b}$ hold necessarily if X_t be well defined, e.g., Filipović (2009, Lemma 10.11) and the arguments above. This completes the proof of Proposition 3.2.

5 Identification of the Bivariate Quadratic Model

The identification of the bivariate quadratic model in Section 3.2 follows from the proof of Proposition 3.2 in Section 4. When X_{1t} is of Class 3, the boundary point 0 is not attainable if and only if $b_1 \geq 1/2$. To prove the necessity of this statement assume that $b_1 < 1/2$. Conditioning on $\beta_{12} X_{2t} < 1/2 - b_1$ for all $t \leq 1$, and using a comparison argument for diffusion processes, see Karatzas and Shreve (1991, Section V.2.C), one can show similarly as in the proof of Proposition 3.2, Class 3,

that $X_{1t} = 0$ for some $t \leq 1$ with non-zero probability. To prove the sufficiency assume that $b_1 \geq 1/2$. The comparison argument for diffusion processes, along with the arguments for Class 3 in the proof of Proposition 3.2, implies that $X_{1t} > 0$ for all t whenever $X_{10} > 0$.

6 Proof of Proposition 3.3

Let $0 \leq n \leq N$. The nth $\mathcal{F}_t$-conditional moment function $f_n(\tau, X_t) = \mathbb{E}_{\mathbb{Q}}\left[X_{t+\tau}^n \mid \mathcal{F}_t\right]$ formally solves the Kolmogorov backward equation

$$\begin{aligned} \frac{\partial}{\partial \tau} f_n(\tau, x) &= \mathcal{A} f_n(\tau, x) \\ f_n(0, x) &= x^n \end{aligned} \tag{6.1}$$

where $\mathcal{A} = (b+\beta x)\frac{\partial}{\partial x} + \frac{1}{2}(a+\alpha x + Ax^2)\frac{\partial^2}{\partial x^2}$ denotes the infinitesimal generator of the quadratic diffusion X_t. We solve (6.1) by the guess $f_n(\tau, x) = \sum_{k=0}^N M_{kn}(\tau) x^k$, for some $(N+1)\times(N+1)$-matrix valued function $M(\tau) = (M_{kn}(\tau))$. Plugging this guess in (6.1), noting that

$$\mathcal{A}x^k = k(k-1)\frac{a}{2}x^{k-2} + k\left(b + (k-1)\frac{\alpha}{2}\right)x^{k-1} + k\left(\beta + (k-1)\frac{A}{2}\right)x^k$$

and matching coefficients in x, we obtain the $N+1$ linear systems of $N+1$ ordinary differential equations

$$\frac{d}{d\tau}\begin{pmatrix} M_{0n}(\tau) \\ M_{1n}(\tau) \\ M_{2n}(\tau) \\ M_{3n}(\tau) \\ \vdots \\ M_{Nn}(\tau) \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & b & 2\frac{a}{2} & 0 & \cdots & 0 \\ 0 & \beta & 2\left(b+\frac{\alpha}{2}\right) & 3\cdot 2\frac{a}{2} & 0 & \vdots \\ 0 & 0 & 2\left(\beta+\frac{A}{2}\right) & 3\left(b+2\frac{\alpha}{2}\right) & \ddots & 0 \\ 0 & 0 & 0 & 3\left(\beta+2\frac{A}{2}\right) & \ddots & N(N-1)\frac{a}{2} \\ \vdots & & & 0 & \ddots & N\left(b+(N-1)\frac{\alpha}{2}\right) \\ 0 & & \cdots & & 0 & N\left(\beta+(N-1)\frac{A}{2}\right) \end{pmatrix}}_{=B} \begin{pmatrix} M_{0n}(\tau) \\ M_{1n}(\tau) \\ M_{2n}(\tau) \\ M_{3n}(\tau) \\ \vdots \\ M_{Nn}(\tau) \end{pmatrix} \tag{6.2}$$

along with the initial condition

$$M_{kn}(0) = \begin{cases} 1, & \text{if } k = n \\ 0, & \text{otherwise.} \end{cases} \tag{6.3}$$

In matrix notation, denote by B the $(N+1)\times(N+1)$ matrix in (6.2), the system (6.2)–(6.3) reads

$$\frac{d}{d\tau}M(\tau) = BM(\tau), \quad M(0) = Id$$

where Id is the identity matrix. Its solution is given by the matrix exponential $M(\tau) = \mathrm{e}^{B\tau}$. It remains to be verified that this provides indeed the nth $\mathcal{F}_t$-conditional moments of $X_{t+\tau}$. Clearly, $f_n(\tau, x) = \sum_{k=0}^n \left(\mathrm{e}^{B\tau}\right)_{kn} x^k$ is a $C^{1,2}$-function whose x-gradient satisfies the polynomial growth condition (1.2).

Hence, Proposition 3.3 follows from the above arguments and Lemma 1.1, noting that $\mathrm{VS}(t,T) = \frac{1}{T-t}\int_0^{T-t} f(\tau, X_t)\,d\tau$.

7 Proof of Theorem 5.1

We first list the technical assumptions that will enable us to prove Theorem 5.1.

Assumption 7.1. *The Radon–Nikodym density* (5.6) *is integrable in the following sense*[58]

$$\mathbb{E}_{\mathbb{Q}}\left[\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\mid_{\mathcal{F}_T}\right)^{-\frac{1}{\eta}}\right] < \infty.$$

Assumption 7.2. *The exponential $\mathbb{Q}$-local martingale*

$$D_t = \exp\left(\frac{1}{\eta}\int_0^t \Lambda(X_s)^\top dW_s - \frac{1}{2\eta^2}\int_0^t \|\Lambda(X_s)\|^2\,ds\right), \quad t\in[0,T] \tag{7.1}$$

is a true martingale. That is, $\mathbb{E}[D_T] = 1$, and we can define the auxiliary equivalent probability measure $\widehat{\mathbb{Q}} \sim \mathbb{Q}$ on $\mathcal{F}_T$ by

$$\frac{d\widehat{\mathbb{Q}}}{d\mathbb{Q}}|_{\mathcal{F}_T} = D_T. \tag{7.2}$$

Assumption 7.3. *The state process X_t is a well-defined diffusion under $\widehat{\mathbb{Q}}$, and the function*

$$h(\tau,x) = \log \mathbb{E}_{\widehat{\mathbb{Q}}}\left[\exp\left(\frac{1}{2\eta}\left(\frac{1}{\eta}-1\right)\int_0^\tau \|\Lambda(X_s)\|^2\,ds\right) \mid X_0 = x\right] \tag{7.3}$$

is of class $C^{1,2}$ on $[0,T]\times\mathcal{X}$. A partial differential equation (PDE) for $H(\tau,x) = \mathrm{e}^{h(\tau,x)}$ is provided in (9.3).

Under Assumptions 5.1 and 7.1 it is well known (e.g., Karatzas and Shreve (1998, Theorem 3.7.6)) that the optimal terminal wealth is given by

$$V_T^* = \left(u'\right)^{-1}\left(\lambda \mathrm{e}^{-rT}\frac{d\mathbb{Q}}{d\mathbb{P}}\mid_{\mathcal{F}_T}\right) \tag{7.4}$$

for some Lagrangian $\lambda = \lambda(V_0)$ such that $\mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-rT}V_T^*\right] = V_0$. Notice that $\left(u'\right)^{-1}(z) = z^{-\frac{1}{\eta}}$. Straight-

[58] In view of (7.5), this is automatically satisfied for relative risk aversion $\eta \geq 1$.

forward manipulations together with (5.6) give

$$\begin{aligned} \mathrm{e}^{-rT}V_T^* &= \lambda^{-\frac{1}{\eta}}\mathrm{e}^{\left(\frac{1}{\eta}-1\right)rT}\exp\left(\frac{1}{\eta}\int_0^T \Lambda(X_s)^\top dW_s - \frac{1}{2\eta}\int_0^T \|\Lambda(X_s)\|^2\,ds\right) \\ &= \lambda^{-\frac{1}{\eta}}\mathrm{e}^{\left(\frac{1}{\eta}-1\right)rT} D_T \exp\left(\frac{1}{2\eta}\left(\frac{1}{\eta}-1\right)\int_0^T \|\Lambda(X_s)\|^2\,ds\right) \end{aligned} \tag{7.5}$$

for the exponential $\mathbb{Q}$-martingale D_t defined in (7.1). Since the discounted optimal wealth process $\mathrm{e}^{-rt}V_t^*$ is a $\mathbb{Q}$-martingale, we thus obtain the equality

$$\begin{aligned} \mathrm{e}^{-rt}V_t^* &= \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-rT}V_T^* \mid \mathcal{F}_t\right] \\ &= \lambda^{-\frac{1}{\eta}}\mathrm{e}^{\left(\frac{1}{\eta}-1\right)rT}\mathbb{E}_{\mathbb{Q}}\left[D_T\exp\left(\frac{1}{2\eta}\left(\frac{1}{\eta}-1\right)\int_0^T \|\Lambda(X_s)\|^2\,ds\right)\mid \mathcal{F}_t\right] \\ &= \lambda^{-\frac{1}{\eta}}\mathrm{e}^{\left(\frac{1}{\eta}-1\right)rT}\mathrm{e}^{h(T-t,X_t)} D_t \exp\left(\frac{1}{2\eta}\left(\frac{1}{\eta}-1\right)\int_0^t \|\Lambda(X_s)\|^2\,ds\right) \\ &= \lambda^{-\frac{1}{\eta}}\mathrm{e}^{\left(\frac{1}{\eta}-1\right)rT}\mathrm{e}^{h(T-t,X_t)}\exp\left(\frac{1}{\eta}\int_0^t \Lambda(X_s)^\top dW_s - \frac{1}{2\eta}\int_0^t \|\Lambda(X_s)\|^2\,ds\right) \end{aligned} \tag{7.6}$$

where the function $h(\tau,x)$ is defined in (7.3).[59] Denote the right hand side of (7.6) by Z_t. Then we obtain via Itô's formula, which is justified by Assumption 7.3,

$$dZ_t = Z_t\left(\frac{1}{\eta}\Lambda(X_t)^\top + \nabla_x h(T-t,X_t)^\top\Sigma(X_t)\right)dW_t. \tag{7.7}$$

On the other hand, we know from (5.3) that

$$d\left(\mathrm{e}^{-rt}V_t^*\right) = \mathrm{e}^{-rt}V_t^*\left(\mathbf{n}_t^\top, w_t\right)\mathcal{G}(t,X_t)\,dW_t. \tag{7.8}$$

Matching coefficients in (7.7) and (7.8) yields the linear system (5.7). By Assumption 5.1, $\mathcal{G}(t,x)^\top$ is surjective. Hence there exists a (possibly non-unique) solution $\mathbf{n}_t^*, w_t^*$. Hence Theorem 5.1 is proved.

[59] Since X_t is a time-homogeneous diffusion under $\widehat{\mathbb{Q}}$, we have

$$\mathbb{E}_{\widehat{\mathbb{Q}}}\left[\exp\left(\frac{1}{2\eta}\left(\frac{1}{\eta}-1\right)\int_t^T \|\Lambda(X_s)\|^2\,ds\right)\mid \mathcal{F}_t\right] = \mathrm{e}^{h(T-t,X_t)}.$$

8 Proof of Corollary 5.1

From the factorization (5.4), it follows that $\mathcal{G}(t,x)$ is injective if and only if $\ker(\mathcal{D}(t,x)\Sigma(x)) \cap \ker\left(\sqrt{g(x)}\,\mathbf{R}(x)^\top\right) = \{\mathbf{0}_{d\times 1}\}$. In view of Assumption 5.1, on one hand this implies that

$$\ker\Sigma(X_t) \cap \ker\left(\sqrt{g(X_t)}\,\mathbf{R}(X_t)^\top\right) = \{\mathbf{0}_{d\times 1}\}, \quad dt\otimes d\mathbb{Q}\text{-a.s.} \tag{8.1}$$

Whence $\dim\ker\Sigma(X_t) \leq 1$ $dt\otimes d\mathbb{Q}$-a.s. in particular, and we conclude that $d \leq n+1$ and $d \leq m+1$. On the other hand it also implies that the maturity date functions $T_i^*(t)$ have to be mutually different for all t.

Now recall the elementary fact from linear algebra that the kernel of $\Sigma(X_t)$ is the orthogonal complement of the image of $\Sigma(X_t)^\top$ in $\mathbb{R}^d$. That is, $\mathbb{R}^d = \ker\Sigma(X_t)\oplus\mathrm{im}\,\Sigma(X_t)^\top$. Taking account of the factorization (5.4), we can rewrite the left hand side of (5.7) as $\Sigma(X_t)^\top\mathcal{D}(t,X_t)^\top\mathbf{n}_t + w_t\sqrt{g(X_t)}\,\mathbf{R}(X_t)$. Since $d > m$, the $m\times d$-matrix $\Sigma(X_t)$ cannot be injective, and there exists a non-zero vector $\mathbf{v} = \mathbf{v}(X_t)$ in $\ker\Sigma(X_t)$. In view of (8.1) it satisfies $\sqrt{g(X_t)}\mathbf{R}(X_t)^\top\mathbf{v} \neq 0$ $dt\otimes d\mathbb{Q}$-a.s. Projecting both sides of (5.7) onto $\mathbf{v}$, we then obtain

$$w_t\sqrt{g(X_t)}\,\mathbf{R}(X_t)^\top\mathbf{v} = \frac{1}{\eta}\Lambda(X_t)^\top\mathbf{v}, \quad dt\otimes d\mathbb{Q}\text{-a.s.} \tag{8.2}$$

Hence the solution $w_t^* = w_t$ of (5.7) is fully determined by the myopic term and does not depend on the choice of the variance swaps. This proves Corollary 5.1.

9 Computation of the Intertemporal Hedging Demand

We now discuss the computation of $\nabla_x h(\tau,x)$ in (5.7). In view of (7.3), if the market price of risk has a constant norm, $\|\Lambda(x)\|^2 \equiv c$, then $h(\tau,x)$ does not depend on x, which implies $\nabla_x h(\tau,x) = 0$, and thus there is no intertemporal hedging demand at all. The same would obviously hold true for the myopic logarithmic utility case $\eta = 1$. In general, $\nabla_x h(\tau,x)$ needs to be computed numerically, e.g., via Taylor expansion. Suppose, for example, that

$$\|\Lambda(x)\|^2 = c + \epsilon^\top\mathbf{P}(x) \tag{9.1}$$

for some constant c, some $\mathbb{R}^k$-valued function $\mathbf{P}(x)$, and some parameter $\epsilon \in \mathbb{R}^k$ with small norm. The first order expansion of $\nabla_x h(\tau, x) = \nabla_x h(\tau, x, \epsilon)$ around $\epsilon = 0$ is[60]

$$\begin{aligned}\nabla_x h(\tau, x, \epsilon) &= \nabla_x h(\tau, x, 0) + \nabla_x \left(\nabla_\epsilon h(\tau, x, \epsilon)|_{\epsilon=0}^\top \epsilon \right) + o(\|\epsilon\|) \\ &= \frac{1}{2\eta} \left(\frac{1}{\eta} - 1 \right) \nabla_x \int_0^\tau \mathbb{E}_{\widehat{\mathbb{Q}}} \left[\epsilon^\top \mathbf{P}(X_s) \mid X_0 = x \right] ds + o(\|\epsilon\|).\end{aligned} \tag{9.2}$$

If the diffusion X_t is quadratic under $\widehat{\mathbb{Q}}$ and $\mathbf{P}(x)$ are polynomials in x, then the conditional moments in the right hand side are available in closed form. Indeed, closed form expressions for $\nabla_x h(\tau, x, \epsilon)$ are available for Taylor expansions of arbitrary order in ϵ.

We next provide a PDE for the function $H(\tau, x) = \mathrm{e}^{h(\tau,x)}$, which could serve as an alternative procedure to compute $\nabla_x h(\tau, x)$. Consider the auxiliary measure $\widehat{\mathbb{Q}}$ defined in (7.2), and denote the respective Girsanov transformed $\widehat{\mathbb{Q}}$-Brownian motion by $d\widehat{W}_t = dW_t - \frac{1}{\eta}\Lambda(X_t)\, dt$. Then the $\widehat{\mathbb{Q}}$-dynamics of X_t reads $dX_t = \left(\mu(X_t) + \frac{1}{\eta}\Sigma(X_t)\Lambda(X_t) \right) dt + \Sigma(X_t)\, d\widehat{W}_t$. The Feynman–Kac theorem thus yields the following linear PDE for $H(\tau, x)$:

$$\begin{aligned}\frac{\partial H(\tau, x)}{\partial \tau} &= \frac{1}{2} \sum_{i,j=1}^m \left(\Sigma(x)\Sigma(x)^\top \right)_{ij} \frac{\partial^2 H(\tau, x)}{\partial x_i \partial x_j} + \left(\mu(x) + \frac{1}{\eta}\Sigma(x)\Lambda(x) \right)^\top \nabla_x H(\tau, x) \\ &\quad + \frac{1}{2\eta} \left(\frac{1}{\eta} - 1 \right) \|\Lambda(x)\|^2 H(\tau, x) \\ H(0, x) &= 1.\end{aligned} \tag{9.3}$$

10 Arguments for the Bivariate Quadratic Model

Following up on Section 5.4, we provide a sketch of the arguments that all assumptions underpinning Theorem 5.1 are satisfied for the bivariate quadratic model in Section 5.3. It follows from Cheridito, Filipović, and Kimmel (2007) that the processes on the right hand side of (5.6) and (7.1) are true martingales for $t \in [0, T]$, and hence define equivalent probability measures $\mathbb{P} \sim \mathbb{Q}$ and $\widehat{\mathbb{Q}} \sim \mathbb{Q}$ on $\mathcal{F}_T$. This asserts validity of Assumption 7.2. Next, we note that the function $g(x)$ as well as the determinant of the 2×2 matrix $\mathcal{D}(t, x)$ are non-zero polynomials in x, for all $t \in [0, T]$, and with smooth t-dependent coefficients. On the other hand, for any $C^{1,2}$-function $\ell(t, x)$ it follows from the occupation times formula that $1_{\{\ell(t,X_t)=0\}} \nabla_x \ell(t, X_t)^\top \Sigma(X_t)\Sigma(X_t)^\top \nabla_x \ell(t, X_t) = 0$, $dt \otimes d\mathbb{Q}$-a.s.; see Revuz and Yor (1994, Corollary (1.6), Chap. VI) and Filipović (2001, Lemma 3.3.1). Since $\Sigma(X_t)\Sigma(X_t)^\top$ is positive definite $dt \otimes d\mathbb{Q}$-a.s., we infer that $\ell(t, X_t) \neq 0$ if $\nabla_x \ell(t, X_t) \neq 0$, $dt \otimes d\mathbb{Q}$-a.s. Applying this to $g(x)$ and the determinant of $\mathcal{D}(t, x)$, we find that $g(X_t) \neq 0$ and $\mathcal{D}(t, X_t)$ is injective, $dt \otimes d\mathbb{Q}$-a.s. Whence validity of Assumption 5.1 follows. Assumption 7.1 holds since we shall consider relative risk aversion $\eta \geq 1$ only. Finally, the $C^{1,2}$-regularity of $h(\tau, x)$ and validity of the PDE (9.3) follows, e.g., from

[60] We omit the effect of ϵ on $\widehat{\mathbb{Q}}$ given by (7.2). This is in particular justified in the bivariate quadratic model, where, in view of (5.8) and (5.9), the $\widehat{\mathbb{Q}}$-law of X_t is invariant with respect to constant shifts of $\|\Lambda(x)\|^2$.

Heath and Schweizer (2000). This asserts Assumption 7.3.

We finally provide an approximation of $\nabla_x h(\tau, x)$ in (5.15). Following up on Appendix 9, we note that the specification (5.11) is of the form (9.1) with $c = \kappa\phi_0$, $\mathbf{P}(x) = \left(\kappa x_1, \kappa x_1^2\right)^\top$, and $\epsilon = (\psi_0, \pi_0)^\top$. The first order Taylor expansion (9.2) then reads

$$\nabla_x h(\tau, x) \approx \frac{1}{2\eta}\left(\frac{1}{\eta} - 1\right) \nabla_x \int_0^\tau \mathbb{E}_{\widehat{\mathbb{Q}}}\left[\kappa\psi_0 X_{1s} + \kappa\pi_0 X_{1s}^2 \mid X_0 = x\right] ds, \tag{10.1}$$

which is available in closed form.

REFERENCES

Ahn, D.-H., R. Dittmar, and A. R. Gallant, 2002, "Quadratic Term Structure Models: Theory and Evidence," *Review of Financial Studies*, 15, 243–288.

Aït-Sahalia, Y., M. Karaman, and L. Mancini, 2012, "The Term Structure of Variance Swaps, Risk Premia and the Expecation Hypothesis," working paper, Princeton University and EPFL.

Aït-Sahalia, Y., and R. Kimmel, 2007, "Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Financial Economics*, 83, 413–452.

Amengual, D., 2009, "The Term Structure of Variance Risk Premia," working paper, Princeton University.

Andrews, D., 1991, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Barone-Adesi, G., R. Engle, and L. Mancini, 2008, "A GARCH Option Pricing Model in Incomplete Markets," *Review of Financial Studies*, 21, 1223–1258.

Bollerslev, T., G. Tauchen, and H. Zhou, 2009, "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 22, 4463–4492.

Bollerslev, T., and V. Todorov, 2011, "Tails, Fears and Risk Premia," *Journal of Finance*, 66, 2165–2211.

Brennan, M. J., and Y. Xia, 2002, "Dynamic Asset Allocation under Inflation," *Journal of Finance*, 57, 1201–1238.

Britten-Jones, M., and A. Neuberger, 2000, "Option Prices, Implied Price Process, and Stochastic Volatility," *Journal of Finance*, 55, 839–866.

Broadie, M., M. Chernov, and M. Johannes, 2007, "Model Specification and Risk Premia: Evidence from Futures Options," *Journal of Finance*, 62, 1453–1490.

Broadie, M., and A. Jain, 2008, "The Effect of Jumps and Discrete Sampling on Volatility and Variance Swaps," *International Journal of Theoretical and Applied Finance*, 11, 761–797.

Buehler, H., 2006, "Consistent Variance Curve Models," *Finance and Stochastics*, 10, 178–203.

Carr, P., and R. Lee, 2010, "Hedging Variance Options on Continuous Semimartingales," *Finance and Stochastics*, 14, 179–207.

Carr, P., and D. Madan, 1998, "Towards a Theory of Volatility Trading," in *Risk Book on Volatility*, ed. by R. Jarrow. Risk, New York, pp. 417–427.

Carr, P., and L. Wu, 2006, "A Tale of Two Indices," *Journal of Derivatives*, 13, 13–29.

——— , 2009, "Variance Risk Premiums," *Review of Financial Studies*, 22, 1311–1341.

Chacko, G., and L. M. Viceira, 2005, "Dynamic Consumption and Portfolio Choice with Stochastic Volatility in Incomplete Markets," *Review of Financial Studies*, 18, 1369–1402.

Chen, L., D. Filipović, and H. V. Poor, 2004, "Quadratic Term Structure Models for Risk-free and Defaultable Rates," *Mathematical Finance*, 14, 515–536.

Cheridito, P., D. Filipović, and R. Kimmel, 2007, "Market Price of Risk Specifications for Affine Models: Theory and Evidence," *Journal of Financial Economics*, 83, 123–170.

Collin-Dufresne, P., and R. Goldstein, 2002, "Do Bonds Span the Fixed-Income Markets? Theory and Evidence for Unspanned Stochastic Volatility," *Journal of Finance*, 57, 1685–1730.

Collin-Dufresne, P., R. Goldstein, and C. Jones, 2008, "Identification of Maximal Affine Term Structure Models," *Journal of Finance*, 63, 743–795.

——— , 2009, "Can Interest Rate Volatility Be Extracted from the Cross Section of Bond Yields?," *Journal of Financial Economics*, 94, 47–66.

Constantinides, G. M., 1992, "A Theory of the Nominal Term Structure of Interest Rates," *Review of Financial Studies*, 5, 531–552.

Dai, Q., and K. J. Singleton, 2000, "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943–1978.

——— , 2003, "Term Structure Modeling in Theory and Reality," *Review of Financial Studies*, 16, 631–678.

Demeterfi, K., E. Derman, M. Kamal, and J. Zou, 1999, "A Guide To Volatility And Variance Swaps," *Journal of Derivatives*, 4, 9–32.

Diebold, F. X., 2012, "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests," working paper, University of Pennsylvania.

Drechsler, I., and A. Yaron, 2011, "What's Vol Got to Do With It," *Review of Financial Studies*, 24, 1–45.

Duarte, J., 2004, "Evaluating an Alternative Risk Preference in Affine Term Structure Models," *Review of Financial Studies*, 17, 379–404.

Duffee, G. R., and R. H. Stanton, 2004, "Estimation of Dynamic Term Structure Models," working paper, University of California-Berkeley.

Duffie, D., D. Filipović, and W. Schachermayer, 2003, "Affine Processes and Applications in Finance," *Annals of Applied Probability*, 13, 984–1053.

Duffie, D., and R. Kan, 1996, "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6, 379–406.

Duffie, D., J. Pan, and K. J. Singleton, 2000, "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.

Dupire, B., 1993, "Model Art," *RISK*, September, 118–120.

Egloff, D., M. Leippold, and L. Wu, 2010, "The Term Structure of Variance Swap Rates and Optimal Variance Swap Investments," *Journal of Financial and Quantitative Analysis*, 45, 1279–1310.

Filipović, D., 2001, *Consistency Problems for Heath-Jarrow-Morton Interest Rate Models*, vol. 1760 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin.

——— , 2002, "Separable Term Structures and the Maximal Degree Problem," *Mathematical Finance*, 12, 341–349.

——— , 2009, *Term-Structure Models*. Springer-Verlag, Berlin.

Filipović, D., and E. Platen, 2009, "Consistent Market Extensions under the Benchmark Approach," *Mathematical Finance*, 19, 41–52.

Fusari, N., and M. T. Gonzalez-Perez, 2012, "Volatility Dynamics and the Term Structure of the Variance Risk Premium," working paper, Northwestern University.

Gatheral, J., 2008, "Developments in Volatility Derivatives Pricing," working paper, Baruch College.

Giacomini, R., and H. White, 2006, "Tests of Conditional Predictive Ability," *Econometrica*, 74, 1545–1578.

Goldstein, R., 2000, "The Term Structure of Interest Rates as a Random Field," *Review of Financial Studies*, 13, 365–384.

Heath, D., and M. Schweizer, 2000, "Martingales Versus PDEs in Finance: An Equivalence Result with Examples," *Journal of Applied Probability*, 37, 947–957.

Jackwerth, J. C., 2000, "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies*, 13, 433–451.

Jarrow, R., Y. Kchia, M. Larsson, and P. Protter, 2013, "Discretely Sampled Variance and Volatility Swaps Versus Their Continuous Approximations," *Finance and Stochastics*, forthcoming.

Jiang, G., and R. Oomen, 2008, "Testing for Jumps When Asset Prices Are Observed With Noise - A "Swap Variance" Approach," *Journal of Econometrics*, 144, 352–370.

Jiang, G., and Y. Tian, 2005, "The Model-Free Implied Volatility and its Information Content," *Review of Financial Studies*, 18, 1305–1342.

Karatzas, I., and S. E. Shreve, 1991, *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York.

——— , 1998, *Methods of Mathematical Finance*. Springer-Verlag, New York.

Kim, T. S., and E. Omberg, 1996, "Dynamic Nonmyopic Portfolio Behavior," *Review of Financial Studies*, 9, 141–161.

Kimmel, R., 2004, "Modeling the Term Structure of Interest Rates: A New Approach," *Journal of*

Financial Economics, 72, 143–183.

Leippold, M., and L. Wu, 2002, "Asset Pricing under the Quadratic Class," *Journal of Financial and Quantitative Analysis*, 37, 271–295.

Liu, J., 2007, "Portfolio Selection in Stochastic Environments," *Review of Financial Studies*, 20, 1–39.

Liu, J., and J. Pan, 2003, "Dynamic Derivative Strategies," *Journal of Financial Economics*, 69, 401–430.

Martin, I., 2013, "Simple Variance Swaps," working paper, Stanford University.

Merton, R. C., 1971, "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory*, 3, 373–413.

——— , 1980, "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.

Meyer, D., and J. Meyer, 2005, "Relative Risk Aversion: What Do We Know?," *Journal of Risk and Uncertainty*, 31, 243–262.

Mueller, P., A. Vedolin, and Y.-M. Yen, 2011, "Bond Variance Risk Premia," working paper, London School of Economics.

Neuberger, A., 1994, "The Log Contract," *Journal of Portfolio Management*, 20, 74–80.

Newey, W., and K. West, 1987, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Revuz, D., and M. Yor, 1994, *Continuous Martingales and Brownian Motion*. Springer-Verlag, Berlin, Germany, second edn.

Sangvinatsos, A., and J. Wachter, 2005, "Does the Failure of the Expectations Hypothesis Matter for Long-Term Investors," *Journal of Finance*, 60, 179–230.

Todorov, V., 2010, "Variance Risk Premium Dynamics: The Role of Jumps," *Review of Financial Studies*, 23, 345–383.

Vuong, Q., 1989, "Likelihood Ratio Tests for Model Selection and Non-Nested Hypothesis," *Econometrica*, 57, 307–333.

Libor Market Model: How to account for the Crisis?

Elise Gourier

I have presented this paper at:

- Brown Bag Seminar of the Institute of Banking and Finance, October 2009, Zürich, Switzerland.
- Annual Swiss Doctoral Workshop in Finance, July 2010, Gerzensee, Switzerland.

Abstract

In this paper we build a new model for Libor rates, which accounts for the stylized effects that appeared during the financial crisis in the dynamics of rates with different tenor structures. Since liquidity and counterparty risks associated to Libor rates depend on the length of the borrowing/lending period, the model is based on a multi-curve approach and reflects the discrepancies which have appeared between rates that used to chase one and another. We define specific dynamics for every Libor rate, depending on its tenor structure, and use Lévy processes as drivers to accommodate for jumps. We provide closed-form expressions in the general setup for the prices of basic interest-rate derivatives. Moreover, we investigate for special cases of Lévy processes, the role of the different parameters on the spread between forward and FRA rates.

1 Introduction

Interest-rate derivatives are by far the largest derivatives market in the world. With a notional amount outstanding of 418 billion USD in December 2008,[61], they represent more than 70% of the total amount outstanding in the global OTC derivatives market. Among the world's 500 largest companies, 88.3% have reported using interest rate derivatives in 2009[62] mostly to manage their cash flows. It is therefore crucial to use accurate and realistic models to analyze and price them.

Different approaches have been developed for modelling interest-rates and inferring prices of derivatives. They can roughly be divided into three categories: short-rate models, instantaneous forward-rate models and Libor market models. Because the latter present the advantage of modelling an observable variable, namely the Libor[63] rate, it has become very popular over the years and is widely used among practitioners.

The standard approach for modelling Libor rates has been introduced by Brace, Gatarek, and Musiela (1997) and Miltersen, Sandmann, and Sondermann (1997) and is based on a single yield curve, constructed using instruments with increasing maturities and without accounting for the tenor structure of the underlying interest rate. However, during the financial crisis significant discrepancies arose between the trajectories of interest rates with different tenor structures. For example, the spread between the one month Libor rate (Libor 1M) and the 30-day Overnight Indexed Swap (OIS)[64] rate, which used to be negligible, has drastically increased since the beginning of the crisis. Because the OIS rate is considered stable and with only little liquidity and counterparty risk, the Libor-OIS spread is an indicator of the banks' perception of risk and is closely watched by financial analysts and economists, as underlined by former US Federal Reserve chairman Alan Greenspan. A larger spread indicates that less money will be available in the markets, leading to higher interest rates and to a decreased confidence of market participants. Further discrepancies have appeared between rates which used to be consistent with one and another, such as quoted FRA rates and forward rates implied by deposits. This phenomenon does not reflect the presence of arbitrage in the markets and can simply be explained by the differences in counterparty and liquidity risk between rates with different tenors. It is therefore fundamental to adapt the pricing algorithm of interest-rate derivatives and to consider the tenor structure of Libor rates as a determinant of their dynamics.

The aim of this paper is to build a realistic model for Libor rates, which accounts for the discrepancies introduced by the financial crisis in the dynamics of rates with different tenor structures. Our model is

[61] According to a report from the Bank of International Settlements published in May 2009

[62] See the 2009 ISDA (International Swaps and Derivatives Association, Inc.) Derivatives Usage Survey, performed on companies of the Fortune Global 500.

[63] The London InterBank Offered Rate is calculated by Thomson Reuters and published by the British Bankers' Association (BBA) each working day. It is used as the basis for settlement of interest rate contracts on many of the world's major future and option exchanges as well as most over the counter and lending transactions. Rates are compiled in ten international currencies and quoted for each currency with 15 maturities, ranging from overnight to 12 months.

[64] OIS are instruments for which one party pays the fixed rate over a predetermined period in return for the average of the overnight Fed funds rate over this period. They do not involve the exchange of a principle and therefore are considered safer than loans. The largest loss incurred is a spread of interest payments accruing over a short period of time.

based on a multi-curve approach and defines a specific dynamic for each Libor rate. Furthermore, our approach integrates the presence of jumps in Libor by using Lévy processes as drivers of the dynamics of rates. These processes indeed offer a satisfying level of flexibility in representing important empirical features of interest rates and have been extensively studied in the context of equity derivatives pricing. Moreover they make it possible to infer closed-form expressions for basic interest-rate derivatives.

Our approach is innovative in the sense that it differentiates for the first time the dynamics of Libor rates according to their tenor structures and integrates jumps through the use of Lévy processes.

Several papers aim to explain the impact of the financial crisis on the modelling of Libor rates. Morini (2008) introduces counterparty risk in the definition of Libor and FRA rates and explains the spread using stochastic default probabilities. However, this approach requires a calibration of a default time to the Libor, which is not as intuitive as in credit risk modelling since the Libor is constructed using a pool of banks and not a single entity. Alternatively, Mercurio (2009) and Bianchetti (2009) model FRA and Libor rates respectively using a multi-curve approach based on standard lognormal Brownian motions with stochastic volatility. Recently, Gefang, Koop, and Potter (2011) used a factor model to represent the spreads between the Libor and the Overnight Index Swap (OIS) rate. Finally, Filipović and Trolle (2013) model interbank risk and decompose its term structure into a default and a liquidity component.

On the other side, the original Libor market model has been revisited a few times, using different classes of processes. Among others, Brigo and Mercurio (2007) give a detailed overview of the existing pre-crisis models. Jarrow, Li, and Zhao (2007) argue that the smile of volatility is also present for options on the Libor and Bates (1996) shows that jumps are able to explain this feature much better than continuous models. Besides, Lim, Ting, and Warachka (2005) argue that jumps are needed in order to accurately price interest rate derivatives. Eberlein and Özkan (2005) are the first to develop a Lévy Libor market model based on time-inhomogeneous Lévy processes. This model is further investigated in Eberlein, Kluge, and Schönbucher (2006), Eberlein and Koval (2006) and Beinhofer, Eberlein, and Janssen (2009). Recently, Keller-Ressel, Papapantoleon, and Teichmann (2012) use affine processes for Libor modelling and Leippold and Stromberg (2013) revisited the Libor market model using time-changed Lévy processes, and estimated their model throughout a time series of caps and swaptions.

The remainder of this paper is divided into six parts. In section 2, we investigate the empirical features of Libor rates with fixed time-to-maturity. In particular, we implement recently developed tests to highlight the presence of jumps. In section 3, we develop a general model using Lévy processes as drivers of the rates and derive closed-form expressions for prices of caplets. In section 4, we consider special cases of Lévy processes and investigate the role of the different parameters on the spread between forward and FRA rates. We show that in the jump diffusion framework, jump parameters have a significant impact on the spread. Finally, we conclude in section 6.

2 Empirical analysis of Libor rates

The standard Libor market model uses a geometric Brownian motion to model Libor rates with a given maturity, disregarding their tenor structure. This raises two issues. First, it assumes that Libor rates are observations of a continuous process with Gaussian log-returns, in particular it excludes the possibility of jumps. Second, it supposes that the tenor structure does not influence the dynamics of the rates.

Although these hypotheses provide tractability in the modelling framework, and in particular allow to derive simple formulas to price derivatives, their ability to capture stylized facts has been questionned in the literature. Therefore it seems essential to first investigate the main empirical features of Libor rates, in order to develop a model which is able to accurately reflect them. The aim of this part is to put into question the two hypotheses presented above for Libor rates with fixed time-to-maturity, using real data[65] from the beginning of 2002 until March 10, 2009. The dataset contains 1825 observations.

2.1 Log-normality of returns

Figure 25 displays the evolution the three-month, six-month and one-year rates over the last eight years. The vertical line indicates the beginning of the crisis, in August 2007. Extreme moves happen on the right side of this line, potentially caused by jumps. They provide a first indication against the hypothesis of log-normality of returns.

[Insert Figure 25 here]

The corresponding daily log-returns are shown on Figure 26. Before the crisis, the volatility seems to decrease with the length of the rate: the one-year Libor rate appears to have a higher volatility than the six-month and the three-month rates. However, we can observe more extreme values for the rates with smaller tenors. This indicates that already before the crisis, using different dynamics for rates with different tenor structures would have been more appropriate.

[Insert Figure 26 here]

Figure 27 represents the kernel smoothed densities of log-returns of Libor rates with different tenor structures. It seems that the distribution becomes less kurtotic when the length between points of the tenor structure increases. The right side of the figure displays the density of the one-year rate and that of the Gaussian density with the same first two moments. The distributions of Libor rates are much heavier-tailed than the normal distribution.

[Insert Figure 27 here]

[65] The data is obtained from the British Bankers' Association.

This observation is confirmed bs the mean-excess plot displayed in Figure 28, which represents the mean of excesses over a threshold as a function of this threshold. Mean-excess plots are usually used as diagnostic plots for the Generalized Pareto Distribution. An upward sloping trend in the mean-excess plot indicates that the distribution is heavy-tailed, with a tail index greater than zero. As a reference, the tail or shape parameter of the normal distribution is zero. Figure 28 clearly shows that the three series of Libor log-returns are upward sloping, with a tail parameter which increases when the lending/borrowing period decreases. The existence of moments higher than one is also put into question here for short lending/borrowing periods.

[Insert Figure 28 here]

Table 17 presents the first four empirical moments of the distributions of the rates, for different tenor structures. Although the moments of order higher than one might not exist and therefore should be manipulated carefully, the summary statistics confirm that the distributions are non-Gaussian, sharp-peaked and heavy-tailed. This feature motivates the use of jump processes.

[Insert Table 17 here]

Finally, Figure 29 displays the realized volatility of Libor log-returns for different tenor structures, calculated as the 100-day moving average volatility:

$$\hat{\sigma}_t^2 = \frac{252}{100} \sum_{j=1}^{100} |r_{t-j}|^2 \tag{2.1}$$

where r_t denotes the log-return at time t.

[Insert Figure 29 here]

The volatility significantly increases after the beginning of the financial crisis, in particular for rates with shorter maturities. Following Cont and Tankov (2003), chapter 7, the realized volatility is however not a good measure of volatility in the presence of heavy tails. Therefore one should not rely too heavily on it, and especially not interpret high variations by the need for stochastic volatility. Therefore we will just consider this graph as an indication that the volatility is likely to be time-varying and dependent on the tenor structure.

2.2 Testing for jumps

As shown in the last section, the densities of Libor log-returns are significantly heavier-tailed than the one of the normal distribution. Empirical studies such as Balduzzi, Elton, and Green (2001) and Fleming and Remonola (1999) argue that extreme variations of prices are likely to be caused by macroeconomic announcements. This would explain why the 1% largest changes in Libor rates are

almost all after September 2007, i.e. after the beginning of the subprime crisis. Furthermore, because large variations might be explained by punctual events, jumps appear to be the perfect tool to use for modelling the behaviour of Libor rates.

We test the presence of jumps by first applying the methodology developed by Aït-Sahalia and Jacod (2009). The test relies on the fact that the estimator $\hat{B}(p, \Delta m)_t$ defined as

$$\hat{B}(p, \Delta n)_t = \sum_{i=1}^{[t/\Delta n]} |\Delta_i^n X|^p. \tag{2.2}$$

X is the process under consideration, Δn denotes the sampling period and

$$\Delta_i^n X = X_{i\Delta n} - X_{(i-1)\Delta n} \tag{2.3}$$

converges toward a quantity which depends on Δn only when $p < 2$, i.e. when the many small increments coming from the continuous part of the process are magnified. When $p > 2$ and large rare jumps overcome the small increments, the ratio of $\hat{B}$ taken at two different sampling frequencies tends to one. Based on that, Aït-Sahalia and Jacod (2009) propose a statistics $\hat{S}(p, k, \Delta n)$, which converges to 1 when the sampling frequency tends to 0 if there are jumps, and to another deterministic value, such as 2, in the absence of jumps.

For each Libor rate we calculate $\hat{S}(p, k, \Delta n)$ with $\Delta n = 5, ..., 1$. In other words, we first calculate the value of the statistics by considering the vector composed of every five observation, then every four etc., until taking the complete vector of rates. The resulting graphs are provided in Figure 30 in the case of the three-month and six-month Libor rates. The grey part of the graph is the region where the null hypothesis that there are no jumps in not rejected, the blue part is the region where the null hypothesis that there are jumps is not rejected. The graphs show that the statistics $\hat{S}(p, k, \Delta n)$ seems to converge toward 1, indicating that it is reasonable to include jumps in the dynamics of Libor rates. Furthermore, we compare the test results using only the pre-crisis data to those using all data. While both test values converge to 1, taking the data of the crisis usually yields a faster convergence. However, this test is designed for high-frequency data and holds asymptotically. Because Libor rates are disclosed on a daily basis, we interpret the test as an indication of the presence of jumps and not as an irrefutable proof.

[Insert Figure 30 here]

To confirm these results, we run the linear and ratio jump tests developed by Barndorff-Nielsen and Shephard (2006). They are respectively based on the difference and ratio between the Realized Bipower Variation (RBV) and Realized Quadratic Variation (RQV), defined as:

$$RBV: \ \{X_{\Delta n}\}_t^{[1,1]} = \sum_{i=2}^{[t/\Delta n]} |\Delta_{i-1}^n X||\Delta_i^n X| \tag{2.4}$$

and

$$RQV: \ [X_{\Delta n}]_t = \sum_{i=1}^{[t/\Delta n]} |\Delta_{i-1}^n X|^2 = \hat{B}(2, \Delta n). \tag{2.5}$$

We separate the data into vectors of 256 datapoints and calculate the linear and ratio jump test values for every period. The null hypothesis is that the data come from a continuous process. If the null hypothesis is satisfied, the statistics should be normally distributed about zero. Figure 31 displays the test values and 10% lower critical values for the Libor 6M rate.

[Insert Figure 31 here]

The test values are most of the time below the critical values, therefore we reject the null hypothesis of continuity. In particular, the difference between the test value and its corresponding critical value is significantly higher in 2008 than the rest of the time. This confirms that a jump process would be more appropriate than a continuous process to model the six-month Libor rate and is in line with the results of the test of Aït-Sahalia and Jacod (2009). The same phenomenon holds for Libor rates with other tenor structures.

2.3 Implications for the modelling approach

The data that we have studied in the last section represents the rates at which banks would borrow funds in the interbank market between time t and time $t+\tau$, where τ characterizes the tenor structure. Let us denote these rates by $F(t, t, t+\tau)$, where the first variable refers to the time at which the decision is made, the second variable marks the beginning of the loan, and the third variable its maturity. This notation will be extended in the next section. Most interest-rate derivatives, such as caps or floors, have as underlying the rate $F(t, T, T+\tau)$, which denotes the rate at which banks would borrow money if they make the decision at time t, and if the borrowing period is $[T, T+\tau]$, for some fixed T. Therefore we aim at modelling this quantity, which is a forward loan with fixed maturity.

The question arises whether we can infer distributional assumptions of forward rates from the empirical study that we have conducted. In other words, do the discontinuities that we have observed in $F(t, t, t+\tau)$ imply the presence of discontinuities in $F(t, T, T+\tau)$, for a given T? Intuitively, the hyperplane $\{t = T\}$ is at the boundary of the domain of definition of the function $F(t, T, T+\tau)$. Assuming that this function does not have jumps implies that the Libor rates can only jump at maturity, therefore jumps would be related to the contract rather than the financial environment. However from an economics perspective, Libor rates are strongly driven by macro-economic events related to the health and credit-worthiness of the financial sector. Therefore we argue that the discontinuities highlighted occur in the t dimension, disregarding whether $t < T$ or $t = T$. Since jumps can cause large deviations from the normal distribution, making the tail of interest rates fatter, the forward rates should be modelled using a distribution which is sufficiently flexible to integrate fat tails.

3 Multi-curve Lévy Libor market model

In the last section we argued that Libor log-returns should be modelled with jump processes and that their dynamics should depend on the tenor structure. These features are incorporated in the model described below.

3.1 Notations

Following Morini (2008), Mercurio (2009) and Bianchetti (2009), empirical discrepancies in the behaviour of Libor rates with different tenor structures motivate the use of a multi-curve approach, where the cash flows of an interest rate derivative product are calculated using the yield curve which corresponds to the tenor structure of the underlying instrument. Similarly, each yield curve should be constructed exclusively with instruments on Libor rates with one single tenor. Such approach breaks down the standard relationship between rates with different tenors, however it implicitely includes a risk component which is is specific to the tenor, such as counterparty and liquidity risk.

In the following we will use the following notations. $\{T_0^f, ..., T_k^f, ..., T_{N_f}^f\}$ denotes the tenor structure associated to the underlying curve $\mathcal{C}_f$, used to determine the future cash flows:

$$\mathcal{C}_f = \{T \to P_f(t,T), T \geq t\} \tag{3.1}$$

where $P_f(t,T)$ is the associated zero-coupon bond price.

There are as many tenor structures as underlying yield curves. To each one of them is associated a Libor rate:

$$F_k^f(t) = \frac{1}{\tau_k^f}\left(\frac{P_f(t,T_{k-1}^f)}{P_f(t,T_k^f)} - 1\right) \quad , \quad t \leq T_{k-1}^f, \tag{3.2}$$

where τ_k^f is the year fraction for the time inverval $[T_{k-1}^f, T_k^f]$ for curve $\mathcal{C}_f$.

For every tenor structure we define the corresponding T_k^f-forward measure $\mathbb{Q}^f_{T_k^f}$, with numeraire $P_f(t,T_k^f)$. Under this measure, F_k^f is a martingale.

Calculating prices of interest rate derivatives involves determining some discount factors. The same curve should be used to discount a payoff on three-month Libor and another payoff on six-month Libor. In other words, pricing formulas should be obtained by using a consistent discounting curve. The construction of this curve is described in Ametrano and Bianchetti (2009). We denote this curve by $\mathcal{C}_d$, and the corresponding T_k^d- forward measure by $\mathbb{Q}^d_{T_k^d}$, with numeraire $P_d(t,T_k^d)$.

We assume that each tenor structure used for forwarding cash flows is included in the discount tenor structure.

3.2 Model

Following Eberlein and Özkan (2005), let us consider an exponential Lévy model and make the two following two assumptions:

(LR.1): For any maturity T_k^f, there is a bounded deterministic function $\lambda(.,T_k^f)$ that represents the volatility of the Libor rate process $F_k^f(.)$.

(LR.2): The initial term structure is given by $F_k^f(0)$ in equation (3.2).

We assume a complete stochastic basis $(\Omega, \mathcal{F}_{T_{N_f}^f}, \mathbb{Q}^f_{T_k^f}, (\mathcal{F}_t)_{0\le t\le T_{N_f}^f})$, where the filtration satisfies the usual conditions. We define the process F_k^f recursively as follows under $\mathbb{Q}^f_{T_k^f}$:

$$F_k^f(t) = F_k^f(0)\exp\left(\int_0^t \lambda(s,T_{k-1}^f)dL_f^{T_k^f}(s)\right) \quad , \quad t \le T_{k-1}^f \tag{3.3}$$

where $L_f^{T_k^f}$ is a Lévy process which has the following canonical representation:

$$L_f^{T_k^f}(t) = bt + c^{\frac{1}{2}}W_f^{T_k^f}(t) + \int_0^t\int_{\mathbb{R}} x(\mu^L - \nu^L)(ds,dx) \quad , \quad t \le T_{k-1}^f. \tag{3.4}$$

In equation (3.4), $W_f^{T_k^f}$ is a standard Brownian motion in $\mathbb{R}$, [66] μ^L is the random measure of jumps[67] of the process $L_f^{T_k^f}$, ν^L is its $\mathbb{Q}_f^{T_k^f}$-compensator. μ^L is related to the Lévy measure F by:

$$\mathbb{E}[\mu^L(.;[0,t]\times A)] = t\cdot F(A) \quad \text{and} \quad \nu^L(ds,dx) = F(dx)ds \tag{3.6}$$

with F satisfying $F(\{0\}) = 0$,

$$\int_0^{T_k^f}\int_{\mathbb{R}} \min(1,x^2)F(dx)ds < \infty \tag{3.7}$$

and

$$\int_0^{T_k^f}\int_{|x|>1} \exp(ux)F(dx)ds < \infty$$

[66] This can be extended to $\mathbb{R}^d$ to get a multi-factor model.

[67] μ^L is defined as

$$\mu^L(w;dt,dx) = \sum_{s>0}\mathbb{I}_{\{\Delta L_s(w)\neq 0\}}\mathcal{E}_{(s,\Delta L_s(w))}(dt,dx) \tag{3.5}$$

Intuitively, $\mu^L(w;[0,t]\times A)$ counts how many jumps of size within A occur for path w from 0 to t.

for $|u| \leq (1+\epsilon)M$ with $M, \epsilon > 0$, and M such that $\sum_{k=1}^{N_f} |\lambda(., T_k^f)| \leq M$. $b \in \mathbb{R}$ is the drift term, and $c \in \mathbb{R}_+^*$.

This model is detailed in the one-curve framework by Eberlein and Özkan (2005). The corresponding correlation structure between F_i^f and F_j^f is described in Beinhofer, Eberlein, and Janssen (2009).

This specification can be extended to a stochastic volatility model by adding another Brownian term in the canonical representation of L. This might provide a significant improvement when calibrating to market data.

3.3 Quanto-type adjustment

Let us consider a payoff on $F_k^f(T_{k-1}^f)$, for example a FRA:

$$\text{Payoff\_FRA}(T_k^f) = F_k^f(T_{k-1}^f) \tag{3.8}$$

or a caplet:

$$\text{Payoff\_caplet}(T_k^f) = \tau_k^f [F_k^f(T_{k-1}^f) - K]^+. \tag{3.9}$$

To price such payoffs, we need to discount them under the discounting measure $\mathbb{Q}^d_{T_k^f}$. While $F_k^f(.)$ is a martingale under $\mathbb{Q}^f_{T_k^f}$, this property is not valid anymore under $\mathbb{Q}^d_{T_k^f}$. Therefore we need to use a change of measure to obtain the expression of $F_k^f(T_{k-1}^f)$ under $\mathbb{Q}^d_{T_k^f}$.

We perform the change of measure using a setup similar to that of quanto options. We refer to Bianchetti (2009) for a description of the analogy. We denote by $X_{fd}(t, T_k^f)$ the ratio of two zero-coupon bonds: $P_f(t, T_k^f)$ associated to curve $\mathcal{C}_f$ and $P_d(t, T_k^f)$ associated to curve $\mathcal{C}_d$:

$$X_{fd}(t, T_k^f) = \frac{P_f(t, T_k^f)}{P_d(t, T_k^f)}. \tag{3.10}$$

We assume that there exists a bounded deterministic function $\sigma(t, T_{k-1}^f)$ for every T_{k-1}^f which represents the volatility of X_{fd}, such that we can write X_{fd} as follows, under $\mathbb{Q}^d_{T_k^f}$:

$$X_{fd}(t, T_k^f) = X_{fd}(0, T_k^f) \exp\left(\int_0^t \sigma(s, T_{k-1}^f) dL_d^{T_k^f}(s) \right) \tag{3.11}$$

with:

$$L_d^{T_k^f}(t) = at + d^{\frac{1}{2}} W_X^{T_k^f}(t) + \int_0^t \int_{\mathbb{R}} x(\mu^L - \bar{\nu}^L)(ds, dx). \tag{3.12}$$

$W_X^{T_k^f}$ is a standard $\mathbb{Q}^d_{T_k^f}$ Brownian motion in $\mathbb{R}$. Finally we define the correlation coefficient ρ_{fX}

such that:

$$d\langle W_f^{T_k^f}, W_X^{T_k^f}\rangle_t = \rho_{fX} dt. \tag{3.13}$$

3.4 Change of measure

Equation (3.3) defines the Libor rate with maturity T_k^f under $\mathbb{Q}^f_{T_k^f}$. Using the martingale property of $F_k^f(.)$, the drift term b is entirely determined:

$$b\int_0^t \lambda(s,T')ds = -\frac{1}{2}c\int_0^t \lambda^2(s,T')ds - \int_0^t\int_{\mathbb{R}} \left(e^{\lambda(s,T')x} - 1 - \lambda(s,T')x\right)\nu^L(ds,dx). \tag{3.14}$$

See proof in Section 1.1 of the appendix. Using equation (3.14), it is equivalent to write $F_k^f(.)$ using a stochastic exponential[68]:

$$F_k^f(t) = F_k^f(0)\mathcal{E}(H_{k-1}^f(t)). \tag{3.16}$$

$H_{k-1}^f(t)$ denotes the exponential transform of $\int_0^t \lambda(s,T_{k-1}^k)dL_f^{T_k^f}(s)$, also called stochastic logarithm of $\frac{F_k^f(t)}{F_k^f(0)}$. $H_{k-1}^f(t)$ is given by:

$$H_{k-1}^f(t) = \int_0^t \lambda(s,T_{k-1}^k)\sqrt{c}\; dW_f^{T_k^f}(s) + \int_0^t\int_{\mathbb{R}} \left(e^{\lambda(s,T_{k-1}^k)x} - 1\right)(\mu^L - \nu^L)(ds,dx). \tag{3.17}$$

The dynamics of $F_k^f(t)$ under $\mathbb{Q}^f_{T_k^f}$ is therefore:

$$\frac{dF_k^f(t)}{F_k^f(t_-)} = \lambda(s,T_{k-1}^k)\sqrt{c}\; dW_f^{T_k^f}(t) + \int_{\mathbb{R}} \left(e^{\lambda(s,T_{k-1}^k)x} - 1\right)(\mu^L - \nu^L)(dt,dx). \tag{3.18}$$

Since $X_{fd}(t,T_k^f)$ is the price of a tradable asset, it is a martingale under the forward measure $\mathbb{Q}^d_{T_k^f}$ with numeraire $P_d(t,T_k^f)$. Similarly to equation 3.14, this entirely defines the drift term a and we can write $X_{fd}(t,T_k^f)$ as follows:

$$X_{fd}(t,T_k^f) = X_{fd}(0,T_k^f)\mathcal{E}(M)_t \tag{3.19}$$

[68]The stochastic exponential $\mathcal{E}(H)$ of a Lévy process H is defined as:

$$\mathcal{E}(H)_t = \exp\left(H_t - \frac{1}{2}\langle H^c\rangle_t\right)\prod_{0\le s\le t}(1+\Delta H_s)\,e^{-\Delta H_s} \tag{3.15}$$

where H^c is the continuous martingale part of the process H.

with:

$$M_t = \int_0^t \sigma(s, T_{k-1}^f)\sqrt{d}\; dW_X^{T_k^f}(s) + \int_0^t \int_{\mathbb{R}} \left(e^{\sigma(s,T_{k-1}^f)x} - 1\right)(\mu^L - \bar{\nu}^L)(ds, dx). \tag{3.20}$$

Equivalently:

$$\frac{dX_{fd}(t, T_k^f)}{X_{fd}(t-, T_k^f)} = \sigma(t, T_{k-1}^f) d^{\frac{1}{2}} dW_X^{T_k^f}(t) + \int_{\mathbb{R}} (e^{\sigma(s,T_{k-1}^f)x} - 1)(\mu^L - \bar{\nu}^L)(ds, dx). \tag{3.21}$$

Furthermore, using the definition of X_{fd}, we can derive the expression of the Radon-Nykodym derivative as follows:

$$\left.\frac{d\mathbb{Q}^f_{T_k^f}}{d\mathbb{Q}^d_{T_k^f}}\right|_{\mathcal{F}_t} = \mathbb{E}\left[\left.\frac{d\mathbb{Q}^f_{T_k^f}}{d\mathbb{Q}^d_{T_k^f}}\right| \mathcal{F}_t\right] = \mathcal{E}(M)_t. \tag{3.22}$$

Applying Girsanov's theorem for semimartingales, as presented in part 12 of Papapantoleon (2005), and originally in Jacod and Shiryaev (2003), with $Y(s,x) = e^{\sigma(t,T_{k-1}^f)x}$, measures $\mathbb{Q}^f_{T_k^f}$ and $\mathbb{Q}^d_{T_k^f}$ are equivalent and we can write under $\mathbb{Q}^d_{T_k^f}$:

$$L_f^{T_k^f}(t) = \bar{b}t + c^{\frac{1}{2}} \tilde{W}_X^{T_k^f}(t) + \int_0^t \int_{\mathbb{R}} x(\mu^L - \bar{\nu^L})(ds, dx). \tag{3.23}$$

The new compensator $\bar{\nu}$ is defined by:

$$\nu^L(ds, dx) = Y(s,x)\bar{\nu}^L(ds, dx) = e^{\sigma(s,T_{k-1}^f)x}\bar{\nu}^L(ds, dx) \tag{3.24}$$

and:

$$\bar{b}t = bt - \rho_{fX}(cd)^{\frac{1}{2}}\sigma(t, T_{k-1}^f) - \int_0^t \int_{\mathbb{R}} x(e^{\sigma(s,T_{k-1}^f)x} - 1)\bar{\nu}^L(ds, dx). \tag{3.25}$$

Y and $\rho_{fX} d^{\frac{1}{2}} \sigma(s, T_{k-1}^f)$ are deterministic but depend on time. Therefore, under measure $\mathbb{Q}^d_{T_k^f}$, $L_f^{T_k^f}$ becomes a process with independent but not stationary increments (additive process).

3.5 Pricing interest-rate derivatives

As done by Eberlein and Özkan (2005), the next step is to use a bilateral Laplace transform to infer the prices of caplets in the special case of purely discontinuous Lévy processes,[69] for example a generalized

[69] i.e., $c = 0$

hyperbolic model.

Following Eberlein and Özkan (2005), we define the following quantities:

$$\begin{aligned} X_{T^f_{k-1}} &= \int_0^{T^f_{k-1}} \lambda(s, T^f_{k-1}) dL^{T^f_k}(s) \\ &= \int_0^{T^f_{k-1}} \lambda(s, T^f_{k-1}) \bar{b} ds + \int_0^{T^f_{k-1}} \int_{\mathbb{R}} x\lambda(s, T^f_{k-1})(\mu^L - \bar{\nu}^L)(ds, dx). \end{aligned} \tag{3.26}$$

As shown in Appendix 1.3, the characteristic function $\chi(u)$ of $X_{T^f_{k-1}}$ is given by:

$$\chi(u) = \exp\left(\int_0^{T^f_{k-1}} \int_{\mathbb{R}} \left[e^{iu\lambda(s,T^f_{k-1})x} - iue^{\sigma(s,T^f_{k-1})x} e^{\lambda(s,T^f_{k-1})x} - (1 - iue^{\sigma(s,T^f_{k-1})x}) \right] \bar{\nu}^L(ds, dx) \right). \tag{3.27}$$

A comparison of equation (3.27) to equation (5.2) of Eberlein and Özkan (2005) shows the effect of the change of measure on the characteristic function of $X_{T^f_{k-1}}$. The term $e^{\sigma(s,T^f_{k-1})x}$ is the single element which arises from the change of measure, it is equal to 1 in the traditional single-curve approach. It also affects the compensator, which is $\bar{\nu}^L(ds, dx)$ instead of $\nu^L(ds, dx)$ in equation (3.27).

We now apply theorem 5.1 of Eberlein and Özkan (2005), which gives the price of a caplet:

Theorem 3.1. *Let $\zeta = -\ln(F^f_k(0))$. Then $F^f_k(T^f_{k-1}) = \exp(-\zeta + X_{T^f_{k-1}})$. For $R < 1$, assume that the moment generating function (mgf) of $X_{T^f_{k-1}}$ exists and that $mgf(R) < \infty$. Let $V(\zeta, \mathcal{K})$ be the time-0 price of a caplet on $F^f_k(T^f_{k-1})$ with strike price $\mathcal{K}$ let $L[\nu_{\mathcal{K}}]$ be the bilateral Laplace transform of $\nu_{\mathcal{K}}$, where $\nu_{\mathcal{K}}$ is defined as follows:*

$$\nu_{\mathcal{K}}(x) = (e^{-x} - \mathcal{K})^+.$$

So we have:

$$L[\nu_{\mathcal{K}}](z) = \int_{-\infty}^{\infty} e^{-zx} \nu_{\mathcal{K}}(x) dx \quad , \quad z = R + iu \in \mathbb{C}, \quad u \in \mathbb{R}.$$

Then:

$$V(\zeta, \mathcal{K}) = (T^f_k - T^f_{k-1}) P_d(0, T^f_k) \frac{e^{\zeta R}}{2\pi} \lim_{M \to \infty} \int_{-M}^{M} e^{iu\zeta} L[\nu_{\mathcal{K}}](R + iu) \chi(iR - u) du \tag{3.28}$$

whenever the right-hand side exists.

A proof of this theorem is provided in Eberlein and Özkan (2005).

4 Analysis of the spread

In the standard single curve approach, cash flows are discounted under the forwarding measure $\mathbb{Q}^f_{T_k^f}$. Therefore the value of a FRA is given, by absence of arbitrage and since $F_k^f(t)$ is a martingale under $\mathbb{Q}^f_{T_k^f}$, by:

$$FRA(t, T_k^f) = \mathbb{E}^{\mathbb{Q}^f_{T_k^f}}[F_k^f(T_{k-1}^f)|\mathcal{F}_t] = F_k^f(t).$$

As shown by Mercurio (2009) in Figure 3, forward rates and FRA rates have been rather aligned until August 2007, therefore this relationship used to hold, but then the two rates started to diverge. This spread is in particular explained by stronger liquidity and counterparty risk on the financial markets.

In this part, we focus on the spread between FRA and forward rates. We look at different models and examine the impact of the parameters on the spread. We consider two different setups: the continuous lognormal setup and the jump-diffusion setup. Both are special cases of the general framework developed in the last section.

4.1 Lognormal setup

In the standard lognormal Libor market model, forward Libor rates evolve according to a geometric Brownian motion under measure $\mathbb{Q}^f_{T_k^f}$:

$$\frac{dF_k^f(t)}{F_k^f(t)} = \lambda(t, T_{k-1}^f) dW_f^{T_k^f}(t). \tag{4.1}$$

For the analogy to section 3.4, see Appendix 2.1.

Under $\mathbb{Q}^d_{T_k^f}$, F_k^f has the following dynamics:

$$\frac{dF_k^f(t)}{F_k^f(t)} = -\lambda(t, T_{k-1}^f)\sigma(t, T_{k-1}^f)\rho_{fX} dt + \lambda(t, T_{k-1}^f) dW_X^{T_k^f}(t). \tag{4.2}$$

We denote the spread between forward and FRA rates at time t by SPREAD_t with:

$$\text{SPREAD}_t = F_k^f(t) - \mathbb{E}^{\mathbb{Q}^d_{T_k^f}}[F_k^f(T_{k-1}^f)|\mathcal{F}_t]. \tag{4.3}$$

As illustrated in Mercurio (2009), the spread should be positive as a consequence of the liquidity and counterparty risk one is exposed to when hedging a FRA position. In other words, forward rates as derived from deposits rates, which are less risky than FRAs.

Equation (2.1) shows that $\log\left(\frac{F_k^f(T_{k-1}^f)}{F_k^f(t)}\right)$ is normally distributed with mean μ and variance γ^2, defined as follows:

$$\mu = -\rho_{fX} \int_0^{T_{k-1}^f} \lambda(t, T_{k-1}^f)\sigma(t, T_{k-1}^f)ds - \frac{1}{2}\int_t^{T_{k-1}} \lambda(s, T_{k-1}^f)ds, \tag{4.4}$$

$$\gamma^2 = \int_t^{T_{k-1}^f} \lambda^2(s, T_{k-1}^f)ds. \tag{4.5}$$

Therefore we have:

$$\text{SPREAD}_t = F_k^f(t)\left[1 - \exp\left(-\rho_{fX}\int_t^{T_{k-1}^f} \lambda(s, T_{k-1}^f)\sigma(s, T_{k-1}^f)ds\right)\right]. \tag{4.6}$$

If we assume that $\lambda(., T_{k-1}^f)$ and $\sigma(., T_{k-1}^f)$ are positive functions, then a positive correlation coefficient ρ_{fX} is needed to preserve the positivity of the spread.

To interpret this result, we need to understand the meaning of the change of measure. Changing the measure gives a different probability to every path of the underlying and hence changes the drift of the process. Under $\mathbb{Q}^f_{T_k^f}$, $F_k^f(t)$ has zero drift while it has drift $-\rho_{fX}\lambda(t, T_{k-1}^f)\sigma(t, T_{k-1}^f)$ under $\mathbb{Q}^d_{T_k^f}$. Indeed, measure $\mathbb{Q}^f_{T_k^f}$ has as numeraire $P_f(t, T_k^f)$, which carries the counterparty and liquidity risk associated to yield curve $\mathcal{C}_f$. On the other hand, $\mathbb{Q}^d_{T_k^f}$ has as a numeraire $P_d(t, T_k^f)$, which is taken from a yield curve with negligible counterparty and liquidity risk, for example the OIS curve. Since $P_f(t, T_k^f)$ is riskier than $P_d(t, T_k^f)$ and both have the same payoff, by absence of arbitrage we must have $P_f(t, T_k^f) < P_d(t, T_k^f)$. Consequently, the price of future cash flows will be smaller when calculated with respect to $P_d(t, T_k^f)$ than when computed with numeraire $P_f(t, T_k^f)$. This justifies why the expected value of the return on $F_k^f(T_{k-1}^f)$ should be smaller under $\mathbb{Q}^d_{T_k^f}$ than under $\mathbb{Q}^f_{T_k^f}$. Because returns under $\mathbb{Q}^f_{T_k^f}$ form a martingale, they have zero mean, hence the drift of returns under $\mathbb{Q}^d_{T_k^f}$ should be negative, which is in line with the positivity of ρ_{fX}.

4.2 Jump diffusion setup

In this part we use a jump diffusion model similar to that of Merton (1976) for the Libor rates, in order to study the impact of the jump parameters on the determination of the spread. $F_k^f(t)$ has the

following dynamics under $\mathbb{Q}^f_{T_k^f}$:

$$F_k^f(t) = F_k^f(0)\exp(L_f^{T_k^f}(t)) \tag{4.7}$$

with

$$L_f^{T_k^f}(t) = bt + c^{\frac{1}{2}} W_f^{T_k^f}(t) + \sum_{i=1}^{N_t} H_i - \gamma\alpha t. \tag{4.8}$$

H_i denote the jump sizes such that $H_i \sim G$ are independent and identically distributed, with $\alpha = \mathbb{E}[H_i]$. In the following we will assume that $H_i \sim \mathcal{N}(\alpha, \beta^2)$. N_t is a Poisson process with intensity γ, i.e. $\gamma t = \mathbb{E}[N_t]$.

For the analogy with the general setup and the technical details of the change of measure, see appendix 2.2. Under $\mathbb{Q}^d_{T_k^f}$, we have:

$$L_f^{T_k^f}(t) = \bar{b}t + c^{\frac{1}{2}} \tilde{W}_X^{T_k^f}(t) + \sum_{i=1}^{\tilde{N}_t} \tilde{H}_i - \gamma e^{\frac{\beta^2}{2}-\alpha}(\alpha - \beta^2)t \tag{4.9}$$

where $\tilde{H}_i$ denote the jump sizes under $\mathbb{Q}_d^{T_k^f}$ such that $\tilde{H}_i$ are i.i.d., with $\tilde{H}_i \sim \mathcal{N}((\alpha - \beta^2), \beta^2)$. $\tilde{N}_t$ is a Poisson process with intensity $\gamma e^{\frac{\beta^2}{2}-\alpha}$. $\bar{b}$ is equal to:

$$\bar{b}t = bt - \rho_{fX} c^{\frac{1}{2}} d^{\frac{1}{2}} t - \gamma[\alpha - e^{\frac{1}{2}\beta^2 - \alpha}(\alpha - \beta^2)]. \tag{4.10}$$

We can show, as it is done in Appendix 2.3, that:

$$\mathbb{E}^{\mathbb{Q}_d^{T_k^f}}[F_k^f(T_{k-1}^f)|\mathcal{F}_t] = F_k^f(t)\exp\left(-\rho_{fX} c^{\frac{1}{2}} d^{\frac{1}{2}} T_{k-1}^f + 2\gamma\left(1 - e^{\frac{\beta^2}{2}}\cosh(\alpha)\right)T_{k-1}^f\right) \tag{4.11}$$

Therefore the spread between forward and FRA rate, as defined above, equals:

$$\text{SPREAD}_t = F_k^f(t)\left[1 - \exp\left(-\rho_{fX} c^{\frac{1}{2}} d^{\frac{1}{2}} T_{k-1}^f + 2\gamma\left(1 - e^{\frac{\beta^2}{2}}\cosh(\alpha)\right)T_{k-1}^f\right)\right] \tag{4.12}$$

To ensure a positive spread, we need the term in the exponential to be negative. The more negative it is, the bigger the spread. When the absolute mean of the jump size gets bigger, the hyperbolic cosine gets bigger too and for β constant, the spread gets bigger. Similarly, when the volatility of jump sizes β gets bigger, so does the spread. These phenomena are magnified by the intensity of jumps γ. Therefore in this model, jumps have the effect of increasing the spread between forward and FRA rates. This is in line with the fact that jumps have been especially present since the beginning of the crisis, i.e. when the spread began to be non zero.

5 Conclusion

In this paper we have constructed a multi-curve Lévy Libor market model which captures the differences in the risks presented by Libor rates with different tenor structures. Based on an empirical analysis, we have highlighted the need to include jumps in the model and proposed a general setup leading to closed-form expressions for caplets. We have considered the special case of a lognormal model and that of a jump diffusion model and we have shown that in the second case, the jump parameters have a significant influence on the spread between FRA rates and forward rates.

	Mean	Variance	Skewness	Kurtosis
OVN	-9.76E-04	0.003	-2.2419	59.5981
1M	-6.54E-04	4.72E-04	1.5844	229.6212
2M	-3.30E-04	2.30E-04	8.1801	335.4483
3M	-1.84E-04	2.07E-04	8.7048	317.4306
4M	-8.86E-05	1.94E-04	5.4244	196.4635
6M	-1.73E-07	2.23E-04	1.817	64.6262
8M	-4.95E-07	2.83E-04	1.3047	38.9044
10M	-9.02E-06	3.56E-04	0.9364	24.8654
1Y	-2.39E-05	4.23E-04	0.7014	18.0026

Table 17: Key statistics of Libor log-returns for different tenor structures. Rates are considered over the period 2002-2009.

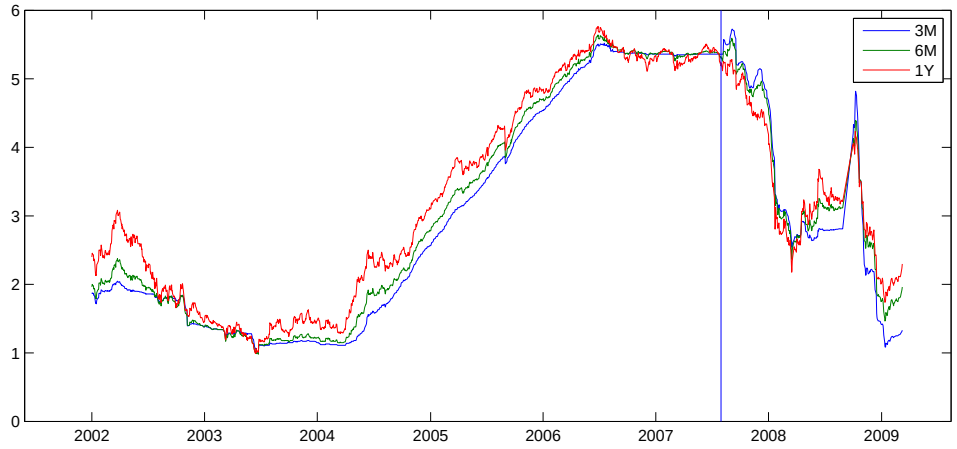


Figure 25: Evolution of USD Libor 3M, Libor 6M and Libor 1Y from 2002 to 2009, in percent. The vertical line indicates the beginning of the financial crisis, in August 2007.

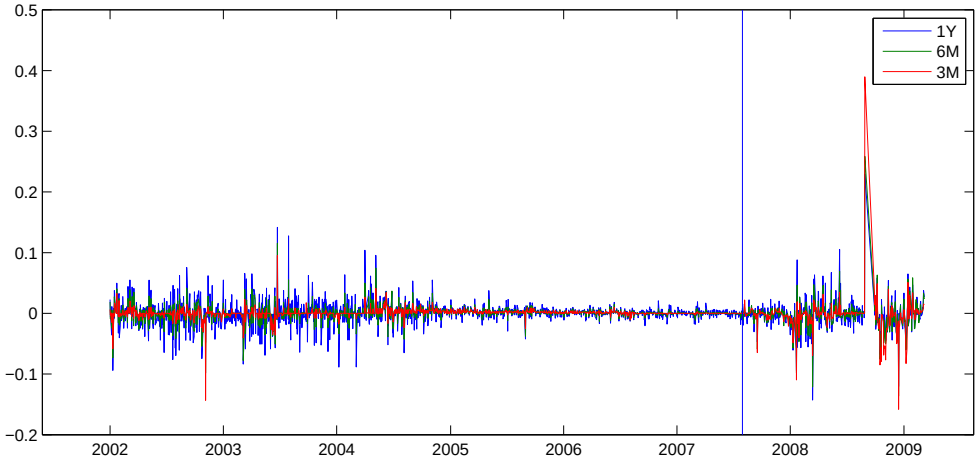


Figure 26: Evolution of USD Libor 3M, Libor 6M and Libor 1Y daily log-returns from 2002 to 2009. The vertical line indicates the beginning of the financial crisis, in August 2007.

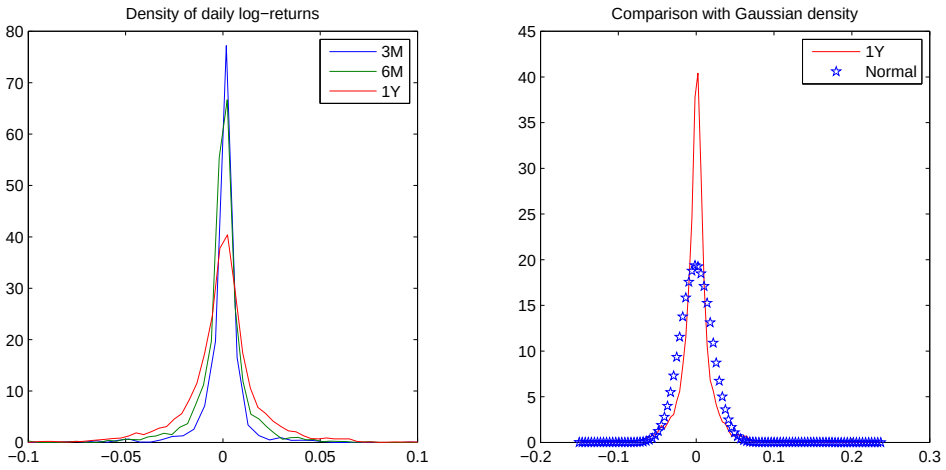


Figure 27: Left graph: Kernel smoothed probability density functions of daily log-returns of Libor rates for tenor structures three months, six months and one-year. Right graph: Comparison of one-year density function with the Gaussian density with same mean and variance.

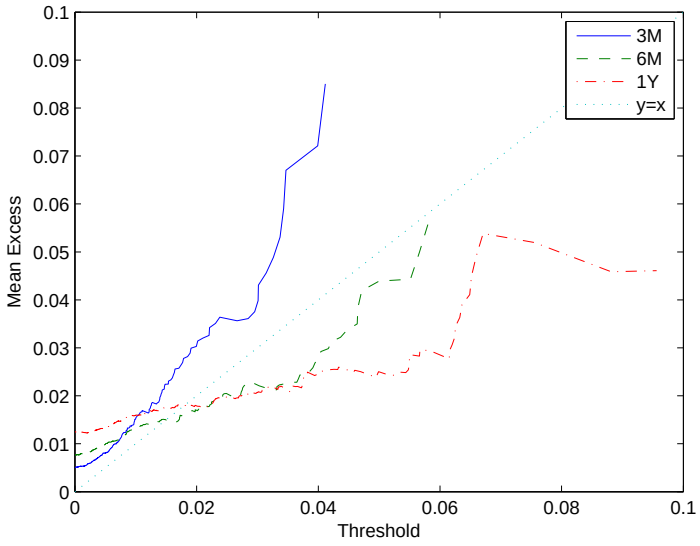


Figure 28: Mean-excess plot of Libor non-negative log-returns for different tenor structures and comparison to the line '$y = x$'.

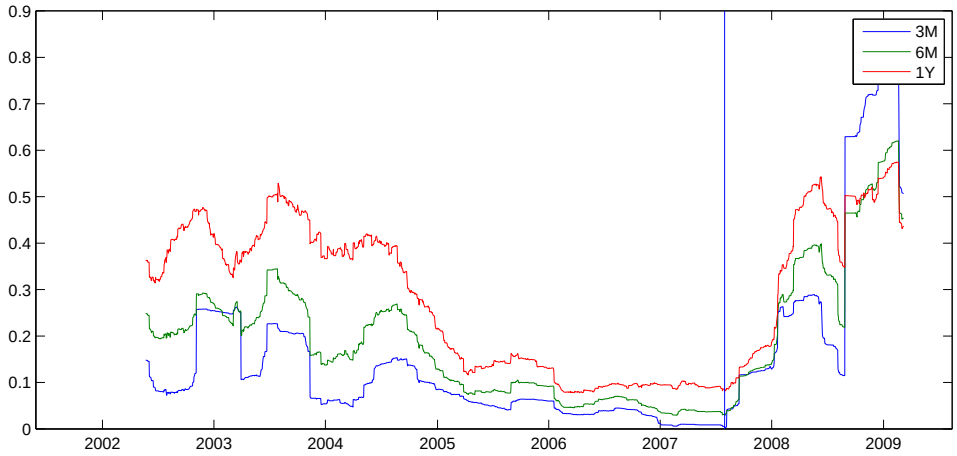


Figure 29: Evolution of the realized (100-day moving average) volatility of the log-returns of Libor rates with different tenor structures from 2002 to 2009. The vertical line indicates the beginning of the financial crisis, in August 2007.

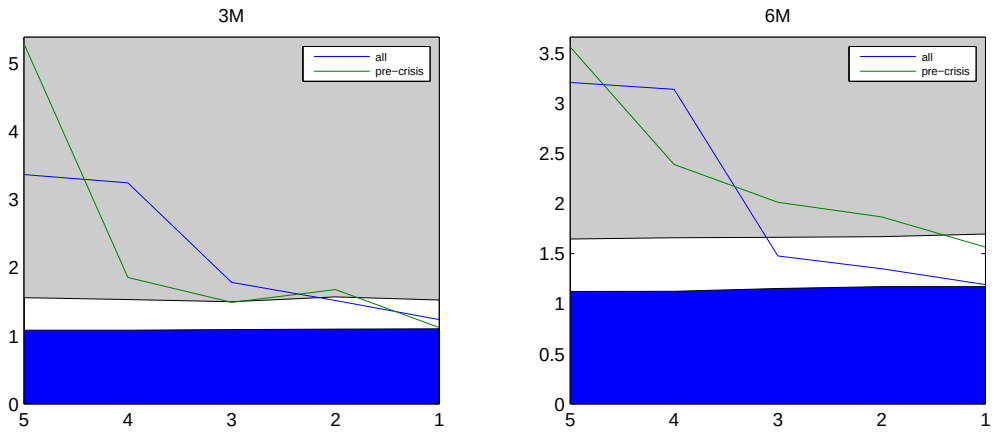


Figure 30: Test statistics $\hat{S}(p, k, \Delta_n)$ for $\Delta_n = 5 : -1 : 1$, using (green curve) only the pre-crisis data and (blue curve) all data. The test value should converge to 1 in the presence of jumps, and otherwise to another deterministic value (such as 2 or 3). The grey part of the graph is the region where the null hypothesis that there are no jumps in not rejected, the blue part is the region where the null hypothesis that there are jumps is not rejected.

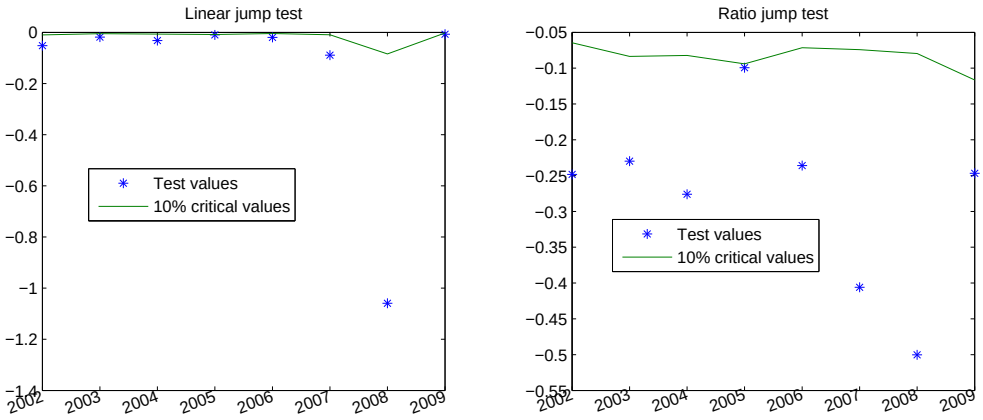


Figure 31: Linear and ratio jump test statistics and their 10% critical values for the Libor 6M rate. The test values should be above the critical values if the null hypothesis that there are no jumps is not rejected.

1 Proofs of section 3

1.1 Drift term of F_k^f under $\mathbb{Q}^f_{T_k^f}$

Here we prove the result given by equation 3.14.

Proof. The martingale condition requires:

$$F_k^f(0) = \mathbb{E}[F_k^f(t)] = \mathbb{E}^{\mathbb{Q}^f_{T_k^f}}[F_k^f(0)\exp\left(\int_0^t \lambda(s, T_{k-1}^k) dL_f^{T_k^f}(s)\right)]$$

$$1 = \mathbb{E}^{\mathbb{Q}^f_{T_k^f}}\left[\exp\left(\int_0^t \lambda(s, T_{k-1}^k) dL_f^{T_k^f}(s)\right)\right]$$

We use the following proposition:

Proposition 1.1 (Eberlein and Raible (1999))**.** *Suppose $f : \mathbb{R}_+ \to \mathbb{C}^d$ is a continuous function such that $|\mathcal{R}(f^i(x))| \leq M$ for all $i \in \{1, ..., d\}$ and $x \in \mathbb{R}_+$, then*

$$\mathbb{E}[\exp\left(\int_0^t f(s) dL_s\right)] = \exp\left(\int_0^t \theta_s(f(s)) ds\right)$$

where θ is defined as follows:

$$\theta_s(z) = \langle z, b_s\rangle + \frac{1}{2}\langle z, c_s z\rangle + \int_{\mathbb{R}^d} \left(e^{\langle z,x\rangle} - 1 - \langle z, x\rangle\right) F_s(dx)$$

for a time-inhomogeneous Lévy process

$$L_t = \int_0^t b_s ds + \int_0^t c_s^{\frac{1}{2}} dW_s + \int_0^t \int_{\mathbb{R}^d} x(\mu^L - \nu)(ds, dx)$$

with characteristics

$$A_t = \int_0^t b_s ds \quad , \quad C_t = \int_0^t c_s ds \quad , \quad \nu(ds, dx) = F_s(dx) ds$$

Applying this proposition we have

$$1 = \mathbb{E}^{\mathbb{Q}^f_{T_k^f}}\left[\exp\left(\int_0^t \lambda(s,T_{k-1}^k)dL_f^{T_k^f}(s)\right)\right] = \exp\left(\int_0^t \theta_s(\lambda(s,T_{k-1}^k))ds\right)$$

$$\Updownarrow$$

$$\int_0^t \theta_s(\lambda(s,T_{k-1}^k))ds = 0$$

$$\Updownarrow$$

$$0 = b\int_0^t \lambda(s,T_{k-1}^k)ds + \frac{1}{2}c\int_0^t \lambda(s,T_{k-1}^k)ds + \int_0^t\int_{\mathbb{R}} \left(e^{\lambda(s,T_{k-1}^k)x} - 1 - \lambda(s,T_{k-1}^k)x\right)\nu^L(dx,ds)$$

□

1.2 Calculation of the stochastic logarithm of $\frac{F_k^f(t)}{F_k^f(0)}$

We prove the result of equation 3.17.

Proof. We use lemma 2.6 of Kallsen and Shiryaev (2002).

Lemma 1.1. *Let X be a real-valued semimartingale with $X_0 = 0$ and let $\tilde{X} = \mathcal{L}(exp(X))$ be its exponential transform (i.e., X is the logarithmic transform of $\tilde{X}$). Then we have:*

(i) $\tilde{X} = X + \frac{1}{2}\langle X^c, X^c\rangle + (e^x - 1 - x) * \mu^X$

(ii) $X = \tilde{X} - \frac{1}{2}\langle \tilde{X}^c, \tilde{X}^c\rangle + (\log(1+x) - x) * \mu^{\tilde{X}}$.

Remark 1.1. *$W * \mu$, for $W = W(w;s,x)$ and the measure $\mu = \mu(w;dt,dx)$, denotes the stochastic integral*

$$W * \mu = \dot{\int_0}\int_{\mathbb{R}} = \dot{\int_0}\int_{\mathbb{R}} W(w;t,x)\mu(w;dt,dx).$$

Furthermore, X^c denotes the continuous martingale part of X. For a Lévy process:

$$L_t^c = c^{\frac{1}{2}}W_t.$$

Applying this lemma, we get:

$$H^f_{k-1}(t) = \int_0^t \lambda(s, T^k_{k-1}) dL_f^{T^f_k}(s) + \frac{1}{2}\langle \left(\int_0^{\cdot} \lambda(s, T^f_{k-1}) dL_f^{T^f_k}(s) \right)^c, \left(\int_0^{\cdot} \lambda(s, T^f_{k-1}) dL_f^{T^f_k}(s) \right)^c \rangle_t + ...$$
$$\int_0^t \int_{\mathbb{R}} \left(e^{\lambda(s, T^f_{k-1})x} - 1 - \lambda(s, T^f_{k-1})x \right) \mu^L(ds, dx)$$

with

$$dL_f^{T^f_k}(t) = bdt + c^{\frac{1}{2}} dW_f^{T^f_k}(t) + \int_{\mathbb{R}} x(\mu^L - \nu^L)(dt, dx).$$

Hence:

$$\begin{aligned}
\int_0^t \lambda(s, T^f_{k-1}) dL_f^{T^f_k}(s) =& b \int_0^t \lambda(s, T^f_{k-1}) ds + c^{\frac{1}{2}} \int_0^t \lambda(s, T^f_{k-1}) dW_f^{T^f_k}(s) + ... \\
& \int_0^t \int_{\mathbb{R}} \lambda(s, T^f_{k-1}) x (\mu^L - \nu^L)(ds, dx) \\
=& -\frac{1}{2} c \int_0^t \lambda^2(s, T^f_{k-1}) ds - \int_0^t \int_{\mathbb{R}} \left(e^{\lambda(s, T')x} - 1 - \lambda(s, T')x \right) \nu^L(ds, dx) + ... \\
& c^{\frac{1}{2}} \int_0^t \lambda(s, T^f_{k-1}) dW_f^{T^f_k}(s) + \int_0^t \int_{\mathbb{R}} \lambda(s, T^f_{k-1}) x (\mu^L - \nu^L)(ds, dx).
\end{aligned}$$

Furthermore:

$$\left(\int_0^{\cdot} \lambda(s, T^f_{k-1}) dL_f^{T^f_k}(s) \right)^c = c^{\frac{1}{2}} \int_0^t \lambda(s, T^f_{k-1}) dW_f^{T^f_k}(s)$$

and

$$\langle \left(\int_0^{\cdot} \lambda(s, T^f_{k-1}) dL_f^{T^f_k}(s) \right)^c, \left(\int_0^{\cdot} \lambda(s, T^f_{k-1}) dL_f^{T^f_k}(s) \right)^c \rangle_t = c \int_0^t \lambda^2(s, T^f_{k-1})) ds.$$

Therefore:

$$
\begin{aligned}
H_F(t,T') = &-\frac{1}{2}c\int_0^t \lambda^2(s,T^f_{k-1})ds - \int_0^t\int_{\mathbb{R}} \left(e^{\lambda(s,T^f_{k-1})x} - 1 - \lambda(s,T^f_{k-1})x\right)\nu^L(ds,dx) + ... \\
&c^{\frac{1}{2}}\int_0^t \lambda(s,T^f_{k-1})dW_f^{T^f_k}(s) + \int_0^t\int_{\mathbb{R}} \lambda(s,T^f_{k-1})x(\mu^L-\nu^L)(ds,dx) + ... \\
&\frac{1}{2}c\int_0^t \lambda^2(s,T^f_{k-1})ds + \int_0^t\int_{\mathbb{R}} \left(e^{\lambda(s,T^f_{k-1})x} - 1 - \lambda(s,T^f_{k-1})x\right)\mu^L(ds,dx) \\
&= c^{\frac{1}{2}}\int_0^t \lambda(s,T^f_{k-1})dW_f^{T^f_k}(s) + \int_0^t\int_{\mathbb{R}} \left(e^{\lambda(s,T')x} - 1\right)(\mu^L-\nu^L)(ds,dx) \qquad (1.1)
\end{aligned}
$$

□

1.3 Calculation of the characteristic function of $X_{T'}$

Applying proposition 1.1 with

$$\bar{\theta}_s(z) = z\bar{b} + \int_{\mathbb{R}} (e^{zx} - 1 - zx)\bar{F}_s(dx)$$

we get the characteristic function of $X_{T^f_{k-1}}$:

$$
\begin{aligned}
\chi(u) &= \mathbb{E}[e^{iuX_{T^f_{k-1}}}] \\
&= \mathbb{E}[e^{iu\int_0^{T^f_{k-1}} \lambda(s,T^f_{k-1})dL^{T^f_k}(s)}] \\
&= \exp\left(\int_0^{T^f_{k-1}} \bar{\theta}_s(iu\lambda(s,T^f_{k-1}))ds\right) \\
&= \exp\left(iu\bar{b}\int_0^{T^f_{k-1}} \lambda(s,T^f_{k-1})ds + \int_0^{T^f_{k-1}}\int_{\mathbb{R}} \left(e^{iu\lambda(s,T')x} - 1 - iu\lambda(s,T^f_{k-1})x\right)\bar{\nu}^L(ds,dx)\right).
\end{aligned}
$$

Furthermore, equation 3.25 gives:

$$\bar{b}t = bt - \int_0^t \int_{\mathbb{R}} x(e^{\sigma(s,T^f_{k-1})x} - 1)\bar{\nu}^L(ds,dx).$$

Therefore:

$$\bar{b}\int_0^{T^f_{k-1}} \lambda(s,T^f_{k-1})ds = b\int_0^{T^f_{k-1}} \lambda(s,T^f_{k-1})ds - \int_0^{T^f_{k-1}} \int_{\mathbb{R}} \lambda(s,T^f_{k-1})x(e^{\sigma(s,T^f_{k-1})x} - 1)\bar{\nu}^L(ds,dx)$$

which gives by equation (3.14):

$$\begin{aligned}
\bar{b}\int_0^{T^f_{k-1}} \lambda(s,T^f_{k-1})ds = &- \int_0^{T^f_{k-1}} \int_{\mathbb{R}} \left(e^{\lambda(s,T^f_{k-1})x} - 1 - \lambda(s,T^f_{k-1})x\right)\nu^L(ds,dx) - ... \\
&\int_0^{T^f_{k-1}} \int_{\mathbb{R}} \lambda(s,T^f_{k-1})x(e^{\sigma(s,T^f_{k-1})x} - 1)\bar{\nu}^L(ds,dx) \\
= &- \int_0^{T^f_{k-1}} \int_{\mathbb{R}} \left(e^{\lambda(s,T^f_{k-1})x} - 1 - \lambda(s,T^f_{k-1})x\right)e^{\sigma(s,T^f_{k-1})x}\bar{\nu}^L(ds,dx) - ... \\
&\int_0^{T^f_{k-1}} \int_{\mathbb{R}} \lambda(s,T^f_{k-1})x(e^{\sigma(s,T^f_{k-1})x} - 1)\bar{\nu}^L(ds,dx).
\end{aligned}$$

Hence $\chi(u)$ can be written as follows:

$$\begin{aligned}
\chi(u) = \exp\Bigg(&-iu \int_0^{T^f_{k-1}} \int_{\mathbb{R}} \left(e^{\lambda(s,T^f_{k-1})x} - 1 - \lambda(s,T^f_{k-1})x\right)e^{\sigma(s,T^f_{k-1})x}\bar{\nu}^L(ds,dx) - ... \\
&iu \int_0^{T^f_{k-1}} \int_{\mathbb{R}} \lambda(s,T^f_{k-1})x(e^{\sigma(s,T^f_{k-1})x} - 1)\bar{\nu}^L(ds,dx) + ... \\
&\int_0^{T^f_{k-1}} \int_{\mathbb{R}} \left(e^{iu\lambda(s,T^f_{k-1})x} - 1 - iu\lambda(s,T^f_{k-1})x\right)\bar{\nu}^L(ds,dx)\Bigg).
\end{aligned}$$

Equivalently,

$$\chi(u) = \exp\left(\int_0^{T_{k-1}^f} \int_{\mathbb{R}} \left[e^{iu\lambda(s,T_{k-1}^f)x} - iue^{\sigma(s,T_{k-1}^f)x} e^{\lambda(s,T_{k-1}^f)x} - (1 - iue^{\sigma(s,T_{k-1}^f)x}) \right] \bar{\nu}^L(ds,dx) \right).$$

2 Proofs of section 4

2.1 Application of the general model to the lognormal setup (4.1)

Let us assume here that the Libor rate corresponding to yield curve $\mathcal{C}_f$ with maturity T_k^f has the following dynamics under $\mathbb{Q}^f_{T_k^f}$:

$$\frac{dF_k^f(t)}{F_k^f(t)} = \lambda(t, T_{k-1}^f) dW_f^{T_k^f}(t).$$

In the general setup of part 3.2, we have $c = 1, \mu^L = \nu^L \equiv 0$. This implies:

$$F_k^f(t) = F_k^f(0) \mathcal{E}\left(\int_0^t \lambda(s, T_{k-1}^f) dW_f^{T_k^f}(s) \right)$$

or equivalently

$$F_k^f(t) = F_k^f(0) \exp\left(\int_0^t \lambda(s, T_{k-1}^f) dL_f^{T_k^f}(s) \right)$$

with

$$L_f^{T_k^f}(t) = bt + W_f^{T_k^f}(t).$$

Using the martingale condition

$$b \int_0^t \lambda(s, T_{k-1}^f) ds = -\frac{1}{2} \int_0^t \lambda^2(s, T_{k-1}^f) ds$$

we get, as expected,

$$F_k^f(t) = F_k^f(0) \exp\left(\int_0^t \lambda^2(s, T_{k-1}^f) dW_f^{T_k^f}(s) - \frac{1}{2} \int_0^t \lambda^2(s, T_{k-1}^f) ds \right).$$

The change of measure is defined by:

$$\left. \frac{d\mathbb{Q}^f_{T_k^f}}{d\mathbb{Q}^d_{T_k^f}} \right| = \mathcal{E}\left(\int_0^t \sigma(s, T') dW_X^{T_k^f}(s) \right).$$

(We have $d = 1$).

Under $\mathbb{Q}^d_{T_k^f}$, the dynamics of $F_k^f(t)$ is therefore defined by:

$$\begin{aligned} \frac{F_k^f(t)}{F_k^f(0)} &= \exp\left(b \int_0^t \lambda(s, T_{k-1}^f) ds - \rho_{fX} \int_0^t \lambda(s, T_{k-1}^f) \sigma(s, T_{k-1}^f) ds + \int_0^t \lambda(s, T_{k-1}^f) d\tilde{W}_X^{T_k^f}(s) \right) \\ &= \exp\left(-\frac{1}{2} \int_0^t \lambda(s, T_{k-1}^f) - \rho_{fX} \int_0^t \lambda(s, T_{k-1}^f) \sigma(s, T_{k-1}^f) ds + \int_0^t \lambda(s, T_{k-1}^f) d\tilde{W}_X^{T_k^f}(s) \right) \end{aligned} \tag{2.1}$$

which gives the dynamics:

$$\frac{dF_k^f(t)}{F_k^f(t)} = -\lambda(s, T_{k-1}^f) \sigma(s, T_{k-1}^f) \rho_{fX} dt + \lambda(s, T_{k-1}^f) d\tilde{W}_X^{T_k^f}(t).$$

2.2 Application of the general model to the jump-diffusion setup (4.2)

$F_k^f(t)$ has the following dynamics under $\mathbb{Q}^f_{T_k^f}$:

$$F_k^f(t) = F_k^f(0) \exp(L_f^{T_k^f}(t))$$

with

$$\begin{aligned} L_f^{T_k^f}(t) &= bt + c^{\frac{1}{2}} W_f^{T_k^f}(t) + \sum_{i=1}^{N_t} H_i - \gamma \alpha t \\ &= bt + c^{\frac{1}{2}} W_f^{T_k^f}(t) + \int_0^t \int_{\mathbb{R}} x(\mu^L - \nu^L)(ds, dx). \end{aligned} \tag{2.2}$$

H_i denote the jump sizes such that $H_i \sim G$ are independent and identically distributed (i.i.d.), with

$\alpha = \mathbb{E}[H_i]$. In the following we will assume that $H_i \sim \mathcal{N}(\alpha, \beta^2)$. N_t is a Poisson process with intensity γ: $\gamma t = \mathbb{E}[N_t]$. In equation 3.3, $\lambda \equiv 1$. Furthermore, the relationships with the random measure of jumps and its compensator are given by:

$$\sum_{i=1}^{N_t} H_i = \int_0^t \int_{\mathbb{R}} x \mu^L(ds, dx)$$

and

$$\mathbb{E}\left[\sum_{i=1}^{N_t} H_i\right] = \gamma \alpha t = \int_0^t \int_{\mathbb{R}} x \nu^L(ds, dx) = t \int_{\mathbb{R}} x \nu^L(dx).$$

Besides, the compensator is related to the law of jumps such that:

$$\nu^L(ds, dx) = \gamma F(dx) ds.$$

Equation 1.1 gives:

$$F_k^f(t) = F_k^f(0) \mathcal{E}(H_F(t, T_{k-1}^f))$$

with

$$H_F(t, T_{k-1}^f) = c^{\frac{1}{2}} W_f^{T_k^f}(t) + \int_0^t \int_{\mathbb{R}} (e^x - 1)(\mu^L - \nu^L)(ds, dx)$$

or equivalently, under $\mathbb{Q}^f_{T_k^f}$:

$$\begin{aligned}
\frac{dF_k^f(t)}{F_k^f(t-)} &= c^{\frac{1}{2}} dW_f^{T_k^f}(t) + \int_{\mathbb{R}} (e^x - 1)(\mu^L - \nu^L)(dt, dx) \\
&= c^{\frac{1}{2}} dW_f^{T_k^f}(t) + d\left(\sum_{i=1}^{N_t} (e^{H_i} - 1)\right) - d\left(\mathbb{E}[\sum_{i=1}^{N_t} (e^{H_i} - 1)]\right) \\
&= c^{\frac{1}{2}} dW_f^{T_k^f}(t) + d\left(\sum_{i=1}^{N_t} (e^{H_i} - 1)\right) - \gamma\left(e^{\alpha + \frac{1}{2}\beta^2} - 1\right) dt.
\end{aligned}$$

Let us now define the Radon-Nykodym density:

$$\frac{d\mathbb{Q}^f_{T^f_k}}{d\mathbb{Q}^d_{T^f_k}} = \mathcal{E}(d^{\frac{1}{2}}W_X^{T^f_k}(.) + \int_0^{\cdot}\int_{\mathbb{R}}(e^x-1)(\mu^L-\nu^L)(ds,dx))_t$$

or, equivalently:

$$\left.\frac{d\mathbb{Q}^f_{T^f_k}}{d\mathbb{Q}^d_{T^f_k}}\right|_t = \exp\left(d^{\frac{1}{2}}W_X^{T^f_k}(t) - \frac{1}{2}d\ t + \int_0^t\int_{\mathbb{R}}x(\mu^L-\nu^L)(ds,dx) - \int_0^t\int_{\mathbb{R}}(e^x-1-x)\nu^L(ds,dx)\right).$$

This change of measure corresponds to the classical Esscher transformation with a fixed price of jump risk equal to 1. The random measure of jumps is a path property, therefore it does not change with the change of measure, which only has an impact on the probability of the paths.

To calculate the new compensator, we use the following proposition:

Proposition 2.1. *Let us consider a compound Poisson process under $\mathbb{Q}^d_T$, with i.i.d. jumps, jump sizes $H_i \sim \mathcal{N}(m, v^2)$ and intensity rate $\lambda(t)$. Assume $\mathbb{Q}^d_T \sim \mathbb{Q}^f_T$ and the density process is defined as:*

$$\left.\frac{d\mathbb{Q}^f_T}{d\mathbb{Q}^d_T}\right|_t = Z_t = \exp\left(\beta c^{\frac{1}{2}}W_t + \int_0^t\int_{\mathbb{R}}\alpha x(\mu^L-\nu^L)(ds,dx) - (\frac{c\beta^2}{2}t + \int_{\mathbb{R}}(e^{\alpha x}-1-\alpha x)\nu(dx))t\right)$$

with $\beta \in \mathbb{R}_{\geq 0}$ and $\alpha \in \mathbb{R}$ constant. Under measure $\mathbb{Q}^f_T$, the compound Poisson process keeps its distributional properties. Only the intensity rate changes to:

$$\bar{\lambda}(t) = \lambda(t)\exp\left(\frac{\alpha(v^2\alpha+2m)}{2}\right)$$

as well as the mean of the jumps:

$$\bar{m} = m + v^2\alpha$$

Proof. Under $\mathbb{Q}^d_T$, the compensator is as follows:

$$\nu(dz,dt) = \lambda(t)F_X(dz)dt$$

where F_X is the distribution of the jumps $\mathcal{N}(m, v^2)$.

Under $\mathbb{Q}^f_T$, Girsanov's theorem gives:

$$\begin{aligned}\bar{\nu}(dz,dt) &= \lambda(t)e^{\alpha z}F_X(dz)dt \\ &= \frac{\lambda(t)}{v\sqrt{2\pi}}\exp\left(-\frac{(z-m)^2}{2v^2}\right)e^{\alpha z}dzdt \\ &= \frac{\lambda(t)e^{\frac{1}{2}\alpha(v^2\alpha+2m)}}{v\sqrt{2\pi}}\exp\left(-\frac{z-(m+v^2\alpha)^2}{2v^2}\right)dzdt \\ &= \bar{\lambda}(t)\bar{F}_X(dz)dt.\end{aligned}$$

By identification, the intensity of jumps under $\mathbb{Q}_T^f$ is given by $\bar{\lambda}(t) = \lambda(t)e^{\frac{1}{2}\alpha(v^2\alpha+2m)}$ and the distribution of jumps is now $\mathcal{N}(m+v^2\alpha, v^2)$

□

We apply this proposition to our case and find that under $\mathbb{Q}^d_{T_k^f}$,

$$\bar{\gamma} = \gamma e^{\frac{\beta^2}{2}-\alpha}$$

$$\bar{\alpha} = \alpha - \beta^2$$

which gives us, under $\mathbb{Q}^d_{T_k^f}$,

$$L_f^{T_k^f}(t) = \bar{b}t + c^{\frac{1}{2}}\tilde{W}_X^T(t) + \sum_{i=1}^{\tilde{N}_t}\tilde{H}_i - \gamma e^{\frac{\beta^2}{2}-\alpha}(\alpha-\beta^2)t \tag{2.3}$$

$\tilde{H}_i$ denote the jump sizes under $\mathbb{Q}^d_{T_k^f}$ such that $\tilde{H}_i$ are i.i.d., with $\tilde{H}_i \sim \mathcal{N}((\alpha-\beta^2),\beta^2)$. $\tilde{N}_t$ is a Poisson process with intensity $\tilde{\gamma}$.

Let us define $\bar{b}'$ such that:

$$\bar{b}t = bt - \rho_{fX}c^{\frac{1}{2}}d^{\frac{1}{2}}t - \bar{b}'t$$

Equations 2.2 and 2.3 give:

$$\bar{b}'t + \sum_{i=1}^{N_t}H_i - \gamma\alpha t = \sum_{i=1}^{\tilde{N}_t}\tilde{H}_i - \gamma e^{\frac{\beta^2}{2}-\alpha}(\alpha-\beta^2)t$$

Since $\sum_{i=1}^{N_t}H_i = \sum_{i=1}^{\tilde{N}_t}\tilde{H}_i$, we have:

$$\bar{b}' = \gamma[\alpha - e^{\frac{1}{2}\beta^2-\alpha}(\alpha - \beta^2)]$$

2.3 Expected value of $F_k^f(T_{k-1}^f)$ under $\mathbb{Q}_{T_k^f}^d$

We have:

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}^d_{T_k^f}}[F_k^f(T_{k-1}^f)|\mathcal{F}_t] =& F_k^f(t)\mathbb{E}^{\mathbb{Q}^d_{T_k^f}}\left[\exp\left(bT_{k-1}^f - \rho_{fX}c^{\frac{1}{2}}d^{\frac{1}{2}}T_{k-1}^f - \gamma[\alpha - e^{\frac{\beta^2}{2}-\alpha}(\alpha-\beta^2)]T_{k-1}^f \cdots\right.\right.\\
&\left.\left. + c^{\frac{1}{2}}\tilde{W}_X^{T_k^f}(T_{k-1}^f) + \sum_{i=1}^{\tilde{N}_{T_{k-1}^f}} \tilde{H}_i - \gamma e^{\frac{\beta^2}{2}-\alpha}(\alpha-\beta^2)T_{k-1}^f\right)\right]\\
=& F_k^f(t)\mathbb{E}^{\mathbb{Q}^d_{T_k^f}}\left[\exp\left(bT_{k-1}^f - \rho_{fX}c^{\frac{1}{2}}d^{\frac{1}{2}}T_{k-1}^f - \gamma\alpha T_{k-1}^f + \frac{1}{2}cT_{k-1}^f + \sum_{i=1}^{\tilde{N}_{T_{k-1}^f}} \tilde{H}_i\right)\right]\\
=& F_k^f(t)\exp\left(bT_{k-1}^f - \rho_{fX}c^{\frac{1}{2}}d^{\frac{1}{2}}T_{k-1}^f - \gamma\alpha T_{k-1}^f + \frac{1}{2}cT_{k-1}^f\right)\mathbb{E}^{\mathbb{Q}^d_{T_k^f}}\left[\exp\left(\sum_{i=1}^{\tilde{N}_{T_{k-1}^f}} \tilde{H}_i\right)\right].
\end{aligned}$$

We use the fact that $\tilde{N}_t$ is $Poisson(\gamma e^{\frac{\beta^2}{2}-\alpha})$ and the moment generating function of a Poisson random variable to get:

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}^d_{T_k^f}}\left[\exp\left(\sum_{i=1}^{\tilde{N}_{T_{k-1}^f}} \tilde{H}_i\right)\right] =& \mathbb{E}^{\mathbb{Q}^d_{T_k^f}}\left[\mathbb{E}^{\mathbb{Q}^d_{T_k^f}}\left[\exp\left(\sum_{i=1}^{n} \tilde{H}_i\right)|\tilde{N}_{T_{k-1}^f} = n\right]\right]\\
=& \mathbb{E}^{\mathbb{Q}^d_{T_k^f}}\left[\exp\left(\tilde{N}_{T_{k-1}^f}(\alpha - \frac{\beta^2}{2})\right)\right]\\
=& \exp\left(\gamma e^{\frac{\beta^2}{2}-\alpha}T_{k-1}^f(e^{\alpha-\frac{\beta^2}{2}} - 1)\right)\\
=& \exp\left(\gamma\left(1 - e^{\frac{\beta^2}{2}-\alpha}\right)T_{k-1}^f\right).
\end{aligned}$$

Finally, we get:

$$\mathbb{E}^{\mathbb{Q}^d_{T^f_k}}[F^f_k(T^f_{k-1})|\mathcal{F}_t] = F^f_k(t)\exp\left(bT^f_{k-1} - \rho_{fX}c^{\frac{1}{2}}d^{\frac{1}{2}}T^f_{k-1} + \frac{1}{2}cT^f_{k-1} + \gamma\left(1-\alpha-e^{\frac{\beta^2}{2}-\alpha}\right)T^f_{k-1}\right).$$

The martingale condition of equation 3.14 further defines b as follows:

$$\begin{aligned} bt &= -\frac{1}{2}ct - \int_0^t\int_{\mathbb{R}}(e^x - 1 - x)\nu^L(ds,dx) \\ &= -\frac{1}{2}ct - \mathbb{E}^{\mathbb{Q}^f_{T^f_k}}[\sum_{i=1}^{N_t}(e^{H_i} - 1 - H_i)] \\ &= -\frac{1}{2}ct - \gamma(e^{\alpha+\frac{\beta^2}{2}} - 1 - \alpha)t. \end{aligned}$$

This gives:

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}^d_{T^f_k}}[F^f_k(T^f_{k-1})|\mathcal{F}_t] =& F^f_k(t)\exp\left(-\rho_{fX}c^{\frac{1}{2}}d^{\frac{1}{2}}T^f_{k-1} + \gamma\left(2 - e^{\frac{\beta^2}{2}-\alpha} - e^{\frac{\beta^2}{2}+\alpha}\right)T^f_{k-1}\right) \\ =& F^f_k(t)\exp\left(-\rho_{fX}c^{\frac{1}{2}}d^{\frac{1}{2}}T^f_{k-1} + 2\gamma\left(1 - e^{\frac{\beta^2}{2}}\cosh(\alpha)\right)T^f_{k-1}\right). \end{aligned}$$

REFERENCES

Aït-Sahalia, Y., and J. Jacod, 2009, "Testing for jumps in a discretely observed process," *The Annals of Statistics*, 37(1), 184—-222.

Ametrano, F. M., and M. Bianchetti, 2009, "Bootstrapping the illiquidity: Multiple yield curves construction for market coherent forward rates estimation," working paper.

Balduzzi, P., E. J. Elton, and T. C. Green, 2001, "Economic news and bond prices: Evidence from the US treasury market," *Journal of Financial and Quantitative Analysis*, 36, 523–543.

Barndorff-Nielsen, O. E., and N. Shephard, 2006, "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4(1), 1—-30.

Bates, D., 1996, "Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options," *Review of Financial Studies*, 9, 69–107.

Beinhofer, M., E. Eberlein, and A. Janssen, 2009, "Correlations in Lévy Interest Rate Models," working paper, Preprint University of Freiburg.

Bianchetti, M., 2009, "Two Curves, One Price: Pricing & Hedging Interest Rate Derivatives Using Different Yield Curves for Discounting and Forwarding," working paper, Banca Intesa Sanpaolo, Milan, Italy.

Brace, A., D. Gatarek, and M. Musiela, 1997, "The market model of interest rate dynamics," *Mathematical Finance*, 7, 127–155.

Brigo, D., and F. Mercurio, 2007, *Interest Rate Models - Theory and Practice*. Springer, Second Edition.

Cont, R., and P. Tankov, 2003, *Financial modelling with Jump Processes*. Chapman & Hall, second edn.

Eberlein, E., W. Kluge, and P. J. Schönbucher, 2006, "The Lévy Libor model with default risk," *Journal of Credit Risk*, 2(2), 3—-42.

Eberlein, E., and N. Koval, 2006, "A cross-currency Lévy market model," *Quantitative Finance*, 6(6), 465—-480.

Eberlein, E., and F. Özkan, 2005, "The Lévy Libor Model," *Finance and Stochastics*, 9, 327—-348.

Eberlein, E., and S. Raible, 1999, "Term structure models driven by general Lévy processes," *Mathematical Finance*, 9, 31–53.

Filipović, D., and A. B. Trolle, 2013, "The Term Structure of Interbank Risk," *Journal of Financial Economics*, Forthcoming.

Fleming, M. J., and E. M. Remonola, 1999, "Price formation and liquidity in the U.S. treasury market: The response to public information," *Journal of Finance*, 54, 1901–1915.

Gefang, D., G. Koop, and S. M. Potter, 2011, "Understanding liquidity and credit risks in the financial crisis," *Journal of Empirical Finance*, 18, 903–914.

Jacod, J., and A. N. Shiryaev, 2003, *Limit Theorems for Stochastic Processes*. Springer, 2nd. edn.

Jarrow, R., H. Li, and F. Zhao, 2007, "Interest Rate Caps "Smile" Too! But Can the LIBOR Market Models Capture It?," *The Journal of Finance*, 62(1).

Kallsen, J., and A. N. Shiryaev, 2002, "The cumulant process and Esscher's change of measure," *Finance and Stochastics*, 6, 397–428.

Keller-Ressel, M., A. Papapantoleon, and J. Teichmann, 2012, "The Affine Libor Models," *Mathematical Finance*, 00, 00.

Leippold, M., and J. Stromberg, 2013, "Time-changed Lévy LIBOR Market Model for the joint Estimation and Pricing of Caps and Swaptions," *Journal of Financial Economics*.

Lim, K. G., C. Ting, and M. Warachka, 2005, "The implied jump risk of LIBOR rates," *Journal of Banking and Finance*, 29, 2503—-2522.

Mercurio, F., 2009, "Interest Rates and The Credit Crunch: New Formulas and Market Models," working paper, QFR, Bloomberg.

Merton, R. C., 1976, "Option pricing when underlying stock returns are discontinuous," *Journal of Financial Economics*, 3, 125–144.

Miltersen, K., K. Sandmann, and D. Sondermann, 1997, "Closed-form solutions for term structure derivaties with lognormal interest rates," *Journal of Finance*, 52, 409–430.

Morini, M., 2008, "The puzzle in the interest rate curve: counterparty risk?," Preprint. Banca IMI, Milan.

Papapantoleon, A., 2005, "An introduction to Lévy processes with applications in finance," .

Part III

Appendix